Volume 69B
Numbers 1 and 2
Jan.–Mar. 1965
Apr.–June 1965

# Mathematics and Mathematical Physics

U.S. DEPARTMENT OF COMMERCE

SECTION B

## JOURNAL OF RESEARCH

NATIONAL BUREAU OF STANDARDS

# Journal of Research

**of the**

**National Bureau of Standards**

**B. Mathematics and Mathematical Physics**

JAN.–MAR. • 1965

APR.–JUNE • 1965

VOLUME • 69B

NUMBERS • 1 and 2

Editor: Chester H. Page

Associate Editors: William H. Pell, John J. Sopka

Publication dates: Mar. 10, June 10, Sept. 10, Dec. 10, 1965


**JOURNAL OF RESEARCH**

The National Bureau of Standards Journal of Research reports research and development in the fields of activity shown at right. Also included from time to time are survey articles on topics closely related to the Bureau's scientific and technical program.

The Journal is published in four separate sections as follows:

**A. Physics and Chemistry**

Contains papers of interest primarily to scientists working in these fields. Issued six times a year. Annual subscription: domestic, $4.00; foreign, $4.75. Single copy 70c.

**B. Mathematics and Mathematical Physics**

Presents studies and compilations designed mainly for the mathematician and the theoretical physicist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75. Single copy 75c.

**C. Engineering and Instrumentation**

Reports research and development results of interest chiefly to the engineer and the applied scientist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75.

**D. Radio Science (formerly Radio Propagation)**

Published in cooperation with U.S. National Committee of URSI. Presents papers in radio propagation, communications, and radio science generally. Issued monthly. Annual subscription: domestic, $9.00; foreign, $11.50. Single copy $1.00.

**The National Bureau of Standards** is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; cooperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information.

**Institute for Basic Standards.** Electricity. Metrology. Heat. Radiation Physics. Mechanics. Applied Mathematics. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radio Standards Laboratory: Radio Standards Physics; Radio Standards Engineering.** Office of Standard Reference Data.

**Institute for Materials Research.** Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.** Office of Standard Reference Materials.

**Central Radio Propagation Laboratory.**** Ionosphere Research and Propagation. Troposphere and Space Communications. Radio Systems. Upper Atmosphere and Space Physics.

**Institute for Applied Technology.** Textiles and Apparel Technology Center. Building Research. Industrial Equipment. Information Technology. Performance Test Development. Instrumentation. Transport Systems. Office of Technical Services. Office of Weights and Measures. Office of Engineering Standards. Office of Industrial Services.

*NBS Group, Joint Institute for Laboratory Astrophysics at the University of Colorado.
**Located at Boulder, Colo.


U.S. DEPARTMENT OF COMMERCE
John T. Connor, Secretary
NATIONAL BUREAU OF STANDARDS
A. V. Astin, Director

Order all publications from the Superintendent of Documents
U.S. Government Printing Office, Washington, D.C., 20402.

Use of funds for printing this publication approved by the Director of the Bureau of the Budget (June 19, 1961).

# The NBS Seminar on Matroids

## August 31–September 11, 1964

The theory and use of matrices, from eigenvalues to linear programming to matrices of integers, has been a major activity at NBS. As part of this program, NBS is interested in bringing to wider attention a recent development, the theory of matroids.

For nine days, between August 31 and September 11, 1964, thirty mathematicians met at NBS in a seminar on matroids. There were about four lectures each day.

W. T. Tutte, Professor at the University of Waterloo, Ontario, Canada, presented the "Lectures on Matroids" which appear here. They are a complete exposition of the theory first published in 1958 and 1959. He presented for the first time an extension of Menger's theorem from networks to matroids, which also appears here.

The other principal speakers at the seminar were:

Louis Auslander, Yeshiva University and U.S. Naval Research Lab.
Henry Crapo, Northeastern University.
R. J. Duffin, Carnegie Institute of Technology.
Jack Edmonds, National Bureau of Standards.
George J. Minty, University of Michigan and Courant Institute.
Gian-Carlo Rota, Massachusetts Institute of Technology.
Horace M. Trent, U.S. Naval Research Laboratory.

Crapo, Edmonds, and Trent have papers published here from the seminar.

The National Bureau of Standards wishes to acknowledge its gratitude to the principal speakers for their impetus and assistance in organizing the seminar, as well as for their outstanding mathematical participation. We also thank the other participants for their extraordinary contribution of talent, thought, good-will, time, and expense. Special thanks are extended for various special contributions to R. Z. Norman, Peter Wegner, John Wilson, John Mather, D. K. Ray-Chaudhuri, and Jack Graver.

Matroids, as an abstraction of matrices, were invented thirty years ago by H. Whitney in a beautiful paper, "The abstract properties of linear dependence", in which he shows that surprisingly many elementary properties are preserved in the abstraction. Until W. T. Tutte's papers in 1958 and 1959, based on his doctoral dissertation in 1948, little more was done with matroids—possibly because it was presumed that there was little to do beyond what is well-known for matrices. Tutte, besides developing fundamentals of general matroids, solved two major problems on matrices in that context.

First, Tutte characterizes totally unimodular matrices. These are matrices such that every minor determinant has value 0, 1, or −1. Such matrices are of great interest in the theory of linear programming. They are essentially equivalent to "regular matroids" and Tutte characterizes regular matroids.

Second, Tutte characterizes those matrices which arise from electrical networks via Kirchhoff's laws—to use his terminology, he characterizes "graphic matroids". Such matrices are naturally of great interest in electrical network design. Graphic matroids are a subclass of regular matroids and the conditions characterizing regular matroids are a subclass of the conditions characterizing graphic matroids. (The proofs of the two characterizations are quite independent.)

The matroid of a matrix is an invariant of the entire group of row vectors generated by adding and subtracting rows of the matrix. Therefore, Tutte treats matroids as arising from these groups—called chain-groups—rather than from matrices. The matroid depends on the particular coordinate representation of the chain-group, and not simply on the "abstract" group. Looked at differently, the matroid of a matrix is determined by the set of abstract points (in the appropriate abstract space) given by the column vectors of the matrix.

Matroids are not generalization for the sake of generalization. Matroids comprise an important area of study about matrices, which also happens to be about other things.

A good reason for relating matroids to matrices is that we all think we know, or think we ought to know, the important facts on matrices. One of the most fundamental aspects of matrices is that they are linear operators or elements of rings. Matroids concern a basically different aspect of matrices—so fundamental, so elementary that it has taken longer to receive attention.

Structures usually presumed too elementary for serious attention are a domain of combinatorial mathematics. Structures which do not, like rings, conform to the "algebra" tradition in discrete mathematics are a domain of combinatorial mathematics. The true context of matroids is combinatorial mathematics.

Louis Auslander
U.S. Naval Research Lab.
Yeshiva University
New York

Peter Bryant
GEC Hirst Research Centre
Wembley, England and
Dept. of Elect. Eng.
University of California, Berkeley

Henry H. Crapo
Northeastern University
Boston, Mass.

R. J. Duffin
Dept. of Mathematics
Carnegie Institute of Technology
Pittsburgh, Pa.

A. Lloyd Dulmage
University of Alberta
Edmonton
Alberta, Canada

Jack Edmonds
Applied Mathematics Division
National Bureau of Standards

D. R. Fulkerson
Rand Corp.
Santa Monica, Calif.

Peter M. Gibson
U.S. Naval Research Lab.
Washington, D.C.

James F. Gimpel
Princeton University
Princeton, N.J.

Alan J. Goldman
Applied Mathematics Division
National Bureau of Standards

Jack E. Graver
Dept. of Mathematics
Dartmouth College
Hanover, N.H.

Irving Katz
Dept. of Mathematics
American University
Washington, D.C.

Alfred Lehman
Water Reed Institute of Research
Washington, D.C.

Joel Levy
Applied Mathematics Division
National Bureau of Standards

John N. Mather
National Bureau of Standards, and
Princeton University
Princeton, N.J.

G. J. Minty
Courant Institute
4 Washington Place
New York, N.Y.

Ronald C. Mullin
Dept. of Mathematics
University of Waterloo
Waterloo, Ontario

Robert Z. Norman
Dept. of Mathematics
Dartmouth College
Hanover, N.H.

Alvin B. Owens
U.S. Naval Research Lab.
Washington, D.C.

D. K. Ray-Chaudhuri
Dept. of Mathematics
IBM Thomas J. Watson Research Center
Yorktown Heights, N.Y.

Fred S. Roberts
Dartmouth College and
Stanford University

Neil Robertson
Dept. of Mathematics
University of Waterloo
Waterloo, Ontario

David Rosenblatt
Information Technology Division
National Bureau of Standards

Gian-Carlo Rota
Dept. of Mathematics, M.I.T.
Cambridge, Mass.

Herbert J. Ryser
Dept. of Mathematics
Syracuse University

Seymour Sherman
Dept. of Mathematics
Indiana University
Bloomington, Ind.

Horace M. Trent
Naval Research Laboratory
Washington, D.C.

W. T. Tutte
Dept. of Mathematics
University of Waterloo
Waterloo, Ontario

Peter Wegner
Dept. of Mathematics
Penn State University
University Park, Pa.

John Wilson
Dept. of Mathematics
University of Waterloo
Waterloo, Ontario

Daniel Younger
Princeton University and
G.E. Research Labs.
Schenectady, New York

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Lectures on Matroids

W. T. Tutte

Professor of Mathematics, University of Waterloo, Ontario, Canada

(November 16, 1964)

This work is based on a series of papers in the Transactions of the American Mathematical Society: A homotopy theorem for matroids, I and II, **88**, 144–174 (1958); and Matroids and graphs, **90**, 572–552 (1959).

These papers set out a theory of matroids, with special emphasis on the conditions for a matroid to represent a graph. The treatment is as rigorous as that in the original papers, but it is hoped that the present work is easier to read, because less condensed. Many theorems originally proved only for "regular matroids" have been generalized to the less restricted class of "binary" ones, and the last part of the work has been improved by the incorporation of a theory of "even" matroids.

## Contents

| | Page |
|---|---|
| 1. Matroids | 1 |
| 1.1. Matroids | 1 |
| 1.2. Chain-groups | 2 |
| 2. Duality | 3 |
| 2.1. Principal forests | 3 |
| 2.2. Dendroids | 3 |
| 2.3. Dendroids of a chain-group | 4 |
| 2.4. Dual chain-groups | 4 |
| 2.5. Bond-matroid of a graph | 5 |
| 2.6. Dual matroids | 6 |
| 3. Structure of matroids | 6 |
| 3.1. Subgraphs and contractions | 6 |
| 3.2. Minors of chain-groups | 7 |
| 3.3. Minors of matroids | 7 |
| 3.4. Connection in matroids | 8 |
| 3.5. Properties of rank | 9 |
| 4. Geometry of matroids | 10 |
| 4.1. Flats | 10 |
| 4.2. Connected flats | 10 |
| 4.3. Linear subclasses | 12 |
| 4.4. Carriers | 13 |
| 5. Specializations | 14 |
| 5.1. Matroids and chain-groups | 14 |
| 5.2. Representative matrices | 15 |
| 5.3. Characterization of binary matroids | 17 |
| 5.4. Regular chain-groups and matroids | 17 |
| 5.5. Some binary matroids which are not regular | 18 |
| 5.6. Graphic and cographic matroids | 19 |
| 6. Homotopy | 20 |
| 6.1. Elementary re-entrant paths | 20 |
| 6.2. Homotopy | 21 |
| 6.3. A lemma | 21 |
| 6.4. Homotopy theorem | 25 |
| 6.5. Special cases | 27 |
| 7. Characterization of regular matroids | 27 |
| 7.1. Some preliminary observations | 27 |
| 7.2. The Fano configuration | 27 |
| 7.3. Heptahedron | 28 |
| 7.4. Condition for a regular matroid | 29 |
| 7.5. Characterization of regular matroid | 32 |
| 8. The matroid and the atom | 32 |
| 8.1. General matroid | 32 |
| 8.2. Binary matroid | 33 |
| 8.3. Kuratowski graphs | 34 |
| 8.4. Skewness | 34 |
| 8.5. Segments determined by bridges | 35 |
| 8.6. Regular matroid | 37 |
| 9. Graphic matroids | 38 |
| 9.1. Bonds | 38 |
| 9.2. Bridges in a bond-matroid | 40 |
| 9.3. Property of odd matroids | 42 |
| 9.4. Characterization of graphic matroids | 46 |

## 1. Matroids

### 1.1. Matroids

Given a graph $G$ we denote the set of vertices by $V(G)$ and the set of edges by $E(G)$.

In general $G$ contains a number of polygons, and each polygon can be specified by its set of edges. Suppose a complete list is made of all the polygons of $G$, each as a set of edges. Then this list can be described as a class of non-null subsets of $E(G)$. We denote it by $P(G)$, and, anticipating a later definition, we call it the *polygon-matroid* of $G$.

The definition of $P(G)$ immediately suggests a difficult graph-theoretical problem. Suppose we are given a class $M$ of non-null subsets of a fixed set $E$. How can we determine whether or not $M$ can be interpreted as the polygon-matroid of a graph? A partial answer can be given at once; it is not difficult to show that the two following conditions are necessary.

I *No member of $M$ is a proper subset of another.*

II *Let $a$ and $b$ be two members of $E$. Let $X$ and $Y$ be members of $M$ such that $a \in X \cap Y$ and $b \in X - Y$. Then there exists a member $Z$ of $M$ such that $a \notin Z$, $b \in Z$ and $Z \subseteq X \cup Y$.*

If $M$ is indeed the polygon-matroid of a graph $G$, then I is immediate from the properties of polygons. To show that II holds in this case we first observe that the polygon with edge-set $X$ contains an arc $L$ which has $b$ as an edge and has only its end-vertices $x$ and $y$ in common with the polygon of edge-set $Y$. But $x$ and $y$ are joined by an arc $L_1$ in the latter polygon such that $a$ is not an edge of $L_1$. Combining $L$ and $L_1$ we obtain a polygon whose edge-set $Z$ has the required properties. (See figure 1A.)

Conditions I and II are not sufficient for $M$ to be of the form $P(G)$. Before going any further it is convenient to lay down the following definition: *a class $M$ of non-null subsets of a set $E$ is a matroid on $E$ if it satisfies conditions I and II.* Thus these two conditions become the axioms of matroid theory.

Our first two theorems are as follows.

1.11 *Let $L$ be a class of non-null subsets of a set $E$. Suppose $L$ satisfies Axiom II. Then if $a \epsilon X \epsilon L$ there is a minimal member $Y$ of $L$ such that $a \epsilon Y \subseteq X$.*

By a "minimal" member of $L$ we mean a member which does not contain another.

We shall denote the number of elements of any finite set $K$ by $|K|$.

To prove our theorem we define $Y$ as a member of $L$ satisfying $a \epsilon Y \subseteq X$ and having the least value of $|Y|$ consistent with this condition. We have to show that $Y$ is a minimal member of $L$.

Suppose not. Then there exists $Z \epsilon L$ such that $Z \subseteq Y - \{a\}$. Choose $b \epsilon Z$. Then by Axiom II, with $a$ and $b$ interchanged, we find that there is a member $Z'$ of $L$ such that $a \epsilon Z' \subseteq Y - \{b\} \subset X$. Since this contradicts the definition of $Y$ the theorem follows.

1.12 *Let $L$ be a class of non-null subsets of a set $E$. Suppose $L$ satisfies Axiom II. Then the minimal members of $L$ constitute a matroid on $E$.*

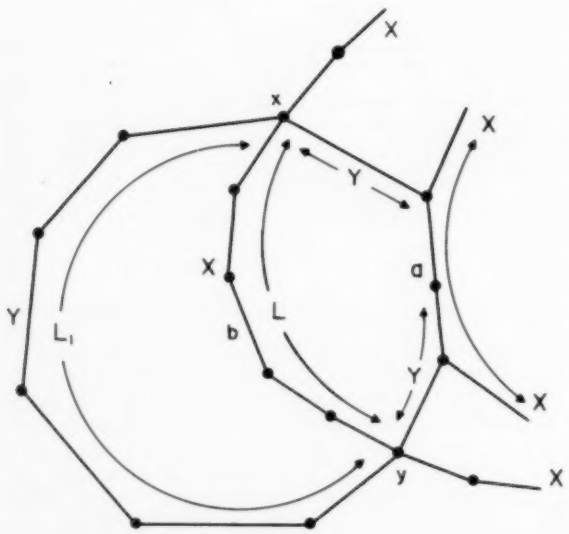



FIGURE 1A

PROOF: Let $M$ be the class of minimal members of $L$. By definition it satisfies Axiom I.

Let $a$ and $b$ be members of $E$, and $X$ and $Y$ members of $M$, such that $a \epsilon X \cap Y$ and $b \epsilon X - Y$. By hypothesis there is a member $W$ of $L$ such that $a \notin W$, $b \epsilon W$ and $W \subseteq X \cup Y$. By 1.1 there exists $Z \epsilon M$ such that $b \epsilon Z \subseteq W$. Hence $a \notin Z$, $b \epsilon Z$ and $Z \subseteq Z \cup Y$. Thus $M$ satisfies Axiom II.

Let $M$ be a matroid on a set $E$. We shall refer to the members of $E$ as the *cells* and to the members of $M$ as the *atoms* of the matroid.

We now describe some further examples of matroids.

## 1.2. Chain-Groups

Let $R$ be a commutative ring with a unit element and no divisors of zero. We define a *chain* on a finite set $E$ over $R$ as a mapping $f$ of $E$ into $R$. Thus each $a \epsilon E$ is associated with a number $f(a)$ in $R$, and this number may be called the coefficient of $a$ in the chain.

The *domain* $||f||$ of $f$ is the set of all members of $E$ with nonzero coefficients in $f$. If $||f||$ is null then $f$ is the zero chain on $E$ over $R$, denoted by the symbol 0.

The *sum* $f+g$ of two chains $f$ and $g$ on $E$ over $R$ is another such chain defined as follows.

$$(f+g)(a) = f(a) + g(a), \qquad a \epsilon E. \tag{1}$$

The product $\lambda f$ of a number $\lambda$ in $R$ and a chain $f$ on $E$ over $R$ is another chain on $E$ over $R$ defined as follows

$$(\lambda f)(a) = \lambda(f(a)), \qquad a \epsilon E. \tag{2}$$

We write $A(E, R)$ for the class of all chains on $E$ over $R$. A chain-group on $E$ over $R$ is defined as any subset $N$ of $A(E, R)$ which is closed under the operations of addition and multiplication by elements of $R$.

Let $N$ be such a chain-group. A chain $f$ of $N$ is called *elementary* if it is nonzero and there is no nonzero $g \epsilon N$ such that $||g|| \subset ||f||$.

A *primitive* chain of $N$ is an elementary chain whose coefficients are restricted to the values 1, 0, and $-1$ in $R$.

We proceed to show that every chain-group has an associated matroid.

1.21 *Let $N$ be a chain-group on $E$ over $R$. Then the class $M$ of the domains of the elementary chains of $N$ is a matroid on $E$.*

*Proof:* Let $L$ be the class of domains of nonzero chains of $N$. Let $f$ and $g$ be two such chains and let $a$ and $b$ be elements of $E$ ("cells" of $N$) such that $a \epsilon ||f|| \cap ||g||$ and $b \epsilon |||f|| - ||g|||$.

We multiply $f$ by $g(a)$ and $g$ by $-f(a)$ and add the resulting chains. We thus obtain a chain $h$ of $N$ such that $a \notin ||h||$, $b \epsilon ||h||$ and $||h|| \subset ||f|| \cup ||g||$. We deduce that $L$ satisfies Axiom II.

But $M$ is the class of minimal members of $L$. Hence $M$ is a matroid on $E$ by 1.2.

We call $M$ the *matroid* of $N$ and denote it by $M(N)$.

A chain group over the ring of residues mod 2 is called *binary*, and one over the ring of integers is called *integral*. A *regular*, or *completely unimodular* chain-group is an integral chain-group in which each elementary chain is a multiple of a primitive chain.

A matroid is called *binary* or *regular* if it is the matroid of a binary or regular chain-group respectively. These two kinds of matroid are of special importance in the theory.

We conclude this chapter by showing how to associate the polygon-matroid of a graph $G$ with a chain-group over $R$.

We first assign an orientation to each edge of $G$ by distinguishing one end as positive and the other as negative. For each $a \epsilon E(G)$ and each $x \epsilon V(G)$ we define an integer $\eta(a, x)$ as follows. If $a$ and $x$ are not incident, or $a$ is a loop, then $\eta(a, x)=0$. Otherwise $\eta(a, x)=1$ or $-1$ according as $x$ is the positive or negative end of $a$.

Chains on $V(G)$ and $E(G)$ are called 0-chains and 1-chains of $G$ respectively.

Let $f$ be a 1-chain of $G$ over $R$. We define a 0-chain $\partial f$ of $G$ over $R$, called the *boundary* of $f$, by the following rule

$$(\partial f)(x)=\sum_{a \epsilon E(G)} \eta(a,x) f(a), \qquad x \epsilon V(G).$$

We call $f$ a *1-cycle* of $G$ if $\partial f=0$. It is clear that the 1-cycles of $G$ over $R$ constitute a chain-group $N$ on $E(G)$ over $R$.

Given a polygon of $G$ we can easily construct a 1-cycle of $G$, with coefficients $1, -1$ and 0 whose domain is the edge-set of the polygon. On the other hand let $f$ be any nonzero 1-cycle over $R$, and let $G_f$ be the subgraph of $G$ made up of the edges of $\|f\|$ and their incident vertices. Then the valency of each vertex of $G_f$ exceeds 1, and therefore $G_f$ contains a polygon. From these observations we have

1.22 *Let $N$ be the chain-group of 1-cycles over $R$ for an oriented graph $G$. Then $M(N)=P(G)$.*

In particular $R$ may be the ring of residues mod 2. Hence

1.23 *The polygon-matroid of a graph $G$ is binary.*

We have seen that the edge-set of any polygon is the domain of a primitive chain of $N$. It readily follows that

1.24 *The polygon-matroid of a graph $G$ is regular.*

## 2. Duality

### 2.1. Principal Forests

A *spanning tree* of a graph $G$ is a subgraph of $G$ which is a tree and which includes all the vertices of $G$. A *principal forest* of $G$ is a subgraph of $G$ whose intersection with each component of $G$ is a spanning tree of that component. For our purposes the main theorem about principal forests is the following.

2.11 *Let $U$ be a subset of $E(G)$. Then $U$ is complementary to the edge-set of a principal forest of $G$ if and only if it has the following properties:*

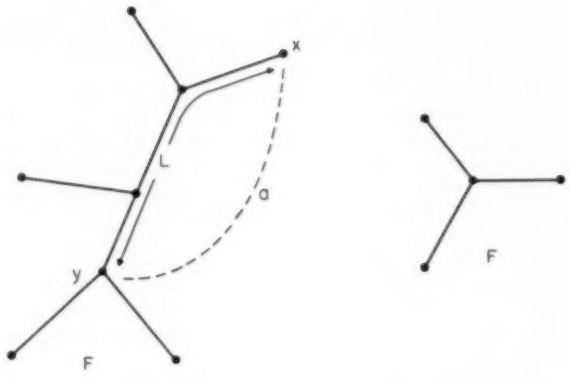



FIGURE 2A

(i) *$U$ meets the edge-set of each polygon of $G$.*

(ii) *No proper subset of $U$ meets the edge-set of every polygon of $G$.*

PROOF: Suppose first that $U$ has properties (i) and (ii). Let $F$ be the spanning subgraph of $G$ such that $E(F)=E(G)-U$. Then $F$ is a forest, by (i). Suppose its intersection with some component $C$ of $G$ is not connected. Then some edge $a \epsilon E(C) \cap U$ has its ends in different components of $F$. But then there can be no polygon of $G$ whose edge-set does not meet $U-\{a\}$, contrary to (ii).

Conversely suppose $U$ is complementary to the edge-set of a principal forest $F$. Then $F$ has no polygon. But if $a$ is an edge of $U$ with ends $x$ and $y$, then either $x=y$ or $x$ and $y$ are joined in $F$ by an arc $L$. (See figure 2A). In the latter case $L$ can be combined with $a$ to form a polygon whose edge-set does not meet $U-\{a\}$. Hence $U$ satisfies (i) and (ii).

### 2.2. Dendroids

We can now make an extension to matroid theory as follows. Let $M$ be a matroid on a set $E$. Then we define a *dendroid* $D$ of $M$ as a subset of $E$ which meets every atom of $M$ and is minimal with respect to this property. Thus

2.21 *The dendroids of $P(G)$ are the complements in $E(G)$ of the edge-sets of the principal forests of $G$.*

Let $D$ be a dendroid of $M$ and let $a$ be a cell of $D$. By the definition of a dendroid there is an atom $X$ of $M$ such that $X \cap D=\{a\}$. Moreover $X$ is unique. For suppose $Y$ is another atom of $M$ such that $Y \cap D=\{a\}$. Then an application of Axiom II shows that there is an atom $Z$ of $M$ not meeting $D$. We denote $X$ by $J(D, a)$.

We now give some general theorems about dendroids of $M$.

2.22 *Let $D$ be a dendroid of $M$. Let $a$ and $b$ be cells of $M$ such that $a \epsilon D$ and $b \notin D$. Write $D'=(D-\{a\}) \cup \{b\}$. Then $D'$ is a dendroid of $M$ if and only if $b \epsilon J(D, a)$.*

PROOF: Suppose $D'$ is a dendroid of $M$. Since it meets $J(D, a)$ we must have $b \epsilon J(D, a)$.

Conversely suppose $b\epsilon J(D, a)$. Then $D'$ meets every atom of $M$, since $D-\{a\}$ meets every such atom other than $J(D, a)$.

We note that $D'-\{b\}$ does not meet $J(D, a)$. Consider any cell $c\epsilon D'-\{b\}$. Then $D'\cap J(D, c)$ is either $\{c\}$ or $\{b, c\}$. In the second alternative an application of Axiom II to $J(D, a)$ and $J(D, c)$ shows that $M$ has an atom $Z$ meeting $D'$ solely in $\{c\}$. We conclude that no proper subset of $D'$ meets every atom of $Z$. The theorem follows.

2.23 *All dendroids of $M$ have the same number of cells.*

PROOF: Let $D_1$ and $D_2$ be distinct dendroids of $M$. Let $D_3$ be a dendroid of $M$ such that $|D_3|=|D_1|$ and $|D_2\cap D_3|$ has the greatest value consistent with this condition.

If possible choose $a\epsilon D_3-D_2$. Then $J(D_3, a)$ meets $D_2$ in a cell $b$ which is not in $D_3$. Since $(D_3-\{a\})\cup\{b\}$ is a dendroid of $M$, by 2.21, the definition of $D_3$ is contradicted.

We deduce that $D_3\subseteq D_2$, and therefore $D_3=D_2$. Hence $|D_1|=|D_2|$, and the theorem follows.

The number of cells in a dendroid of $M$ is the *rank* $r(M)$ of $M$. In the trivial case in which $M$ has no atoms there is just one dendroid, the null subset of $E$, and $r(M)=0$. In every other case $r(M)>0$.

2.24 *Let $X$ be an atom of $M$. Then there is a dendroid $D$ of $M$ and a cell $a$ of $D$ such that $X=J(D, a)$.*

PROOF: Let $D$ be a dendroid of $M$ with the least possible value of $|D\cap X|$. Suppose $D\cap X$ has two distinct cells $a$ and $b$. Then $J(D, b)$ has a cell $c$ not in $X$, by Axiom I. But $(D-\{b\})\cup\{c\}$ is a dendroid of $M$, by 2.3, which is contrary to the definition of $D$. We conclude that $D\cap X$ consists of a single cell $a$, so that $X=J(D, a)$.

COROLLARY. *If $x\epsilon X$ there is a dendroid $D'$ of $M$ such that $X=J(D', x)$.* (2.22 and 2.24).

2.25 *A matroid $M$ on a given set $E$ is uniquely determined by its dendroids.*

PROOF: If $D$ is any dendroid of $M$ and $a\epsilon D$, then $J(D, a)$ is uniquely determined by the other dendroids, by 2.22. Hence $M$ is completely determined by its dendroids, by 2.24.

## 2.3. Dendroids of a Chain-Group

The concept of a dendroid can be applied to a chain-group $N$ on $E$ over a commutative ring $R$ of the type described in chapter I. We define a *dendroid* of $N$ as a minimal subset $D$ of $E$ meeting the domain of every nonzero chain of $N$. Since every such domain contains that of an elementary chain the dendroids of $N$ are identical with those of the corresponding matroid $M(N)$.

Let $D$ be a dendroid of $N$. For each $a\epsilon D$ we can find a chain $f_a$ of $N$ such that $\|f_a\|$ is the atom $J(D, a)$ of $M(N)$. We refer to the set of chains $f_a$, one for each $a\epsilon D$, as a *dendroid-basis* of $N$. Its *weight* is the product

$$\prod_{a\epsilon D} f_a(a),$$

which is necessarily nonzero.

2.31 *Let $B=\{f_a|a\epsilon D\}$ be a dendroid-basis of $N$ of weight $w$. Let $K$ be any non-zero chain of $N$. Then $wK$ can be expressed as a linear combination of members of $B$, with multipliers in $R$.*

PROOF: By adding suitable multiples of members of $B$ to $wK$ we can reduce the coefficient of each $a\epsilon D$ to zero. The resulting chain is zero, since its domain does not meet $D$.

COROLLARY: *If $\|K\|$ meets $D$ in a single cell $a$, then $\|K\|=\|f_a\|$, and therefore $K$ is an elementary chain of $N$.*

## 2.4. Dual Chain-Groups

We note that the rank of $M(N)$ is also the rank of $N$, in the sense of the theory of linear dependence, by 2.31. Much simplification results if we can arrange that $f_a(a)=1$ for each $a\epsilon D$. This can be done when $R$ is a field and also when $N$ is regular. In the latter case $f_a$ is taken to be a primitive chain of $N$. In these cases the weight of the dendroid-basis takes the value 1. The dendroid-basis then becomes a true basis of $N$, by 2.31.

Returning to the general case we define two chains $f$ and $g$ on $E$ over $R$ as *orthogonal* if

$$\sum_{a\epsilon E} f(a)g(a)=0.$$

Let $N^*$ be the class of all chains $h$ on $E$ over $R$ such that $h$ is orthogonal to every member of $N$. It is easily verified that $N^*$ is a chain-group on $E$ over $R$. We call it the *dual* chain-group of $N$. It is clear from the definition that

2.41 $$N\subseteq N^{**}$$

2.42 *Let $D$ be a dendroid of $N$ and let $B=\{f_a|a\epsilon D\}$ be a corresponding dendroid-basis of $N$.*

*For each $b\epsilon E-D$ let a chain $g_b$ on $E$ be defined as follows: $g_b(b)$ is the weight $w(B)$ of $B$, and $g_b(x)=0$ for each $x\epsilon(E-D)-\{b\}$. On the other hand if $a\epsilon D$, then*

$$g_b(a)=-f_a(b)\prod_{c\epsilon D-\{a\}} f_c(c).$$

*Then $E-D$ is a dendroid of $N^*$, and the set $B^*=\{g_b|b\epsilon E-D\}$ is a corresponding dendroid-basis of $N^*$.*

PROOF: For each $g_b\epsilon B^*$ it is readily verified that $g_b$ is orthogonal to each $f_a\epsilon B$. Hence $g_b$ is orthogonal to each chain of $N$, by 2.31 and the fact that $R$ has no divisors of zero. Hence $g_b\epsilon N^*$.

A chain on $E$ over $R$ whose domain is a non-null subset of $D$ cannot be orthogonal to every member of $B$. Hence $E-D$ meets the domain of every nonzero chain of $N^*$. Considering the chains $g_b$ we see that no proper subset of $E-D$ has this property. Hence $E-D$ is a dendroid of $N^*$. By 2.31, Corollary, $B^*$ is a corresponding dendroid-basis.

2.43 *The dendroids of $N^*$ are the complements in $E$ of the dendroids of $N$.*

PROOF: Any dendroid of $N$ is also a dendroid of $N^{**}$, by two applications of 2.42. Hence $N$ and $N^{**}$ have equal ranks.

Let $E-U$ be any dendroid of $N^*$. Then $U$ is a dendroid of $N^{**}$, by 2.42. It contains a dendroid $U'$ of $N$, by 2.41, and $U'=U$ by the result of the preceding paragraph. Thus $U$ is a dendroid of $N$. Combining this result with 2.42 we establish the theorem.

2.44 $$M(N^{**})=M(N),$$

by 2.25 and 2.43.

2.45 *Any dendroid-basis of $N$ is a dendroid-basis of $N^{**}$*, by 2.41, 2.44 and 2.31, Corollary.

2.46 *If $N$ has a dendroid-basis of weight 1, then $N^{**}=N$*, by 2.31 and 2.45.

2.47 *If $N$ is regular, then $N^*$ is regular.*

PROOF: We may suppose the chains $f_a$ of 2.42 to be primitive. Then the chains $g_b$ have coefficients restricted to the values 0, 1 and $-1$. Since they belong to a dendroid-basis of $N^*$ they are elementary and therefore primitive chains of $N^*$. Using 2.24 we deduce that $N^*$ is regular.

## 2.5. Bond-Matroid of a Graph

Suppose now that $N$ is the group of 1-cycles on a graph $G$ over $R$. Then the dendroids of $N^*$ are the edge-sets of the principal forests of $G$, by 1.22, 2.21, and 2.43. We proceed to interpret $N^*$ in terms of the structure of $G$.

Let $g$ be any 0-chain on $G$ over $R$. Then the *coboundary* $\delta g$ of $g$ is a 1-chain on $G$ over $R$ defined as follows

$$(\delta g)(a)=\sum_{x\epsilon V(G)} \eta(a, x)g(x), \qquad a\epsilon E(G).$$

It is clear that the coboundaries of the 0-chains on $G$ over $R$ are the elements of a chain-group $N_1$ on $E(G)$ over $R$.

2.51 *Every coboundary of $G$ over $R$ is orthogonal to every 1-cycle of $G$ over $R$.*

PROOF: Let $k$ be a 1-cycle and $f$ the coboundary of a 0-chain $g$. Then

$$\sum_{a\epsilon E(G)} k(a)f(a)=\sum_{a\epsilon E(G)}\sum_{x\epsilon V(G)} \eta(a, x)k(a)g(x)=\sum_{x\epsilon V(G)} g(x)h(x),$$

where $h=\partial k$. Hence

$$\sum_{a\epsilon E(G)} k(a)f(a)=0.$$

As a corollary we note that $N_1\subseteq N^*$.

Now let $F$ be a dendroid of $N^*$, the edge-set of a principal forest of $G$.

Consider any $a\epsilon F$, belonging to a component $C_a$ of $G$. The intersection of the principal forest with $C_a$ is a tree $T$ having $a$ as an edge. When $a$ is deleted $T$ is decomposed into two components $T_1$ and $T_2$, which we may suppose to contain the positive and negative ends of $a$ respectively. (See figure 2B.)

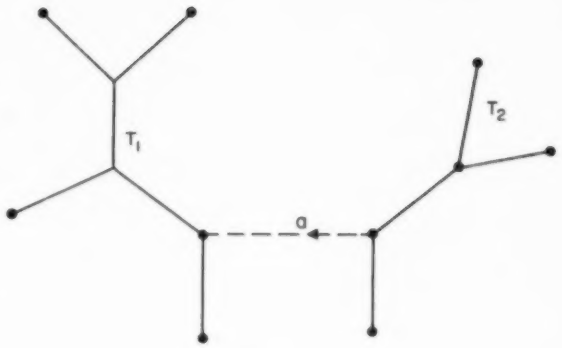



FIGURE 2B

Let $g_a$ be the 0-chain over $R$ in which the vertices of $T_1$ have coefficient 1 and all other vertices coefficient 0. Consider the coboundary $f_a=\delta g_a$. Clearly $f_a(a)=1$ and the other coefficients in $f_a$ are restricted to the values $1, -1$ and 0. Moreover $F\cap\|f_a\|=\{a\}$. The last result implies, since $N_1\subseteq N^*$, that $f_a$ is an elementary chain of $N^*$, by 2.31, Corollary.

The coboundaries $f_a$, $a\epsilon F$, thus constitute a dendroid-basis of $N^*$. Using 2.31 we obtain

2.52 *$N^*$ is the chain-group of coboundaries of $G$ over $R$*

We refer to $M(N^*)$ as the *bond-matroid* $B(G)$ of $G$. We note that in the above proof $\|f_a\|$ is not dependent on the particular ring $R$ used. Hence, by 2.24, Corollary, $B(G)$ is independent of $R$.

Since $R$ may be the ring of residues mod 2 we have

2.53 *$B(G)$ is binary.*

If $R$ is the ring of integers we observe that the chains $f_a$ are all primitive. Hence, by 2.24, Corollary, we have

2.54 *$B(G)$ is regular.*

We refer to the atoms of $B(G)$ as the *bonds* of $G$. As an exercise in graph theory we can show that a non-null subset $K$ of $E(G)$ is a bond of $G$ if and only if it satisfies the following condition: the graph $H$ obtained from $G$ by deleting the edges of $K$ has two components $C_1$ and $C_2$ such that each edge of $K$ has one end in each.

We have adopted a nomenclature which treats $B(G)$ as more fundamental then $P(G)$. Thus principal forests of $G$ correspond directly to "dendroids" of $B(G)$. We also describe a matroid as *graphic* if it can be represented as the bond-matroid of a graph, and as *cographic* if it can be represented as a polygon-matroid.

There is a theory of *dual graphs* on the 2-sphere, but we shall comment only briefly about it. Two graphs $G$ and $H$ may be defined as dual if there is a $1-1$ mapping of $E(G)$ onto $E(H)$ which transforms 1-cycles into coboundaries, and coboundaries into 1-cycles. It can be shown that a graph is *planar*, that is realizable in the 2-sphere, if and only if it has a dual graph. Using this theory we may assert that a graph $G$ is planar if and only if its polygon-matroid is graphic. Let us

therefore define a *planar* matroid as one which is both graphic and cographic. Such a matroid corresponds to a pair of dual planar graphs.

### 2.6. Dual Matroids

There is a theory of duality for general matroids which is closely analogous to that given for chain-groups in section 2.4.

Two subsets $S$ and $T$ of a finite set $E$ are called *orthogonal* if $|S \cap T| \neq 1$. Given a matroid $M$ on $E$ we denote the class of all non-null subsets of $E$ which are orthogonal to every atom of $M$, by $L(M)$.

2.61 *If* $(X, Y) \in L(M)$, $a \in X \cap Y$ *and* $b \in X - Y$, *then there exists* $Z \in L(M)$ *such that* $b \in Z \subseteq (X \cup Y) - \{a\}$.

PROOF: Assume the theorem false for some $X, Y, a, b$. We construct a sequence $(a_0, a_1, \ldots, a_k)$ of cells of $X \cup Y$, and a sequence $(T_1, T_2, \ldots, T_k)$ of atoms of $M$ as follows. First we put $a_0 = a$. If we have determined the cells $a_i$ as far as $a_r$, and $b$ is not among them we take $T_{r+1}$ to be any atom of $M$ such that

$$T_{r+1} \cap \{(X \cup Y) - \{a_0, \ldots, a_r\}\}$$

has just one cell, $c$ say. This is possible since $(X \cup Y) - \{a_0, \ldots, a_r\}$ is not in $L(M)$, by assumption. We then write $c = a_{r+1}$. The construction terminates with $a_k = b$.

There exists $U \in M$ such that $b \in U \cap (X \cup Y) \subseteq \{a_0, \ldots, a_k\}$. For example we may take $U = T_k$. For each such $U$ let $p(U)$ be the greatest integer $j$ less than $k$ such that $a_j \in U$. Such an integer exists since $|U \cap X| \neq 1$.

Suppose $p(U) > 0$. We apply Axiom II to $U$ and $T_{p(U)}$, thus establishing the existence of an atom $U'$ of $M$ such that

$$b \in U' \subseteq (U \cup T_{p(U)} - \{a_{p(U)}\})$$

But then $b \in U' \cap (X \cup Y) \subseteq \{a_0, \ldots, a_k\}$, and $p(U') < p(U)$.

We can therefore choose $U$ so that $p(U) = 0$. But then $Y \cap U = \{a_0\} = \{a\}$, contrary to hypothesis. The theorem follows.

We denote the class of minimal members of $L(M)$ by $M^*$. Now 2.61 asserts that $L(M)$ satisfies Axiom II. Hence,

2.62 $M^*$ *is a matroid on* $E$.

by 1.12. We call $M$ the *dual matroid* of $M$. From the definition of orthogonality we see that

$$M \subseteq M^{**}. \tag{1}$$

If $D$ is a dendroid of $M$ and $b \in E - D$ we define $K(D, b)$ as the subset of $E$ consisting of $b$ and each $a \in D$ such that $b \in J(D, a)$.

2.63 $K(D, b) \in M^*$.

PROOF: Assume $K(D, b)$ is not in $L(M)$. Then there exists an atom $X$ of $M$ such that $|X \cap K(D, b)| = 1$. Choose such an $X$ so that $|D \cap X|$ has the least possible value, and denote the common cell of $X$ and $K(D, b)$ by $c$.

Suppose $D \cap X$ has a cell $e$ not in $K(D, b)$. Applying Axiom II to $X$ and $J(D, e)$ we find that $M$ has an atom $Y$ such that $c \in Y \subseteq (X \cup J(D, e)) - \{e\}$. But then $Y \cap K(D, b) = \{c\}$, since $b \notin J(D, e)$. Moreover $|D \cap Y| < |D \cap X|$. This is contrary to the definition of $X$.

We deduce that $D \cap X \subseteq X \cap K(D, b) = \{c\}$. Hence, since $D$ is a dendroid of $M$, $D \cap X = \{c\}$ and therefore $X = J(D, c)$. But $c \in K(D, b)$ and therefore $b \in J(D, c)$. Hence $X \cap K(D, b)$ includes two distinct cells $b$ and $c$, which is contrary to the definition of $X$. Hence $K(D, b)$ is in fact in $L(M)$.

There is an atom $U$ of $M^*$ such that $b \in U \subseteq K(D, b)$, by 1.11. But if $x \in K(D, b) - U$, then $U$ is not orthogonal to $J(D, x)$. Hence $K(D, b) = U \in M^*$.

COROLLARY: $E - D$ *is a dendroid of* $M^*$, *and* $K(D, b)$ *is the atom* $J(E - D, b)$ *of* $M^*$.

2.64 *The dendroids of* $M^*$ *are the complements in* $E$ *of the dendroids of* $M$.

PROOF: The proof is analogous to that of 2.43. We note that any dendroid of $M$ is a dendroid of $M^{**}$, by 2.63, Corollary. Hence $r(M^{**}) = r(M)$.

Let $E - U$ be any dendroid of $M^*$. Then $U$ is a dendroid of $M^{**}$. It contains a dendroid $U'$ of $M$, by (1). But then $U' = U$ by the equality of rank.

Combining this result with the Corollary to 2.63 we establish the theorem.

2.65

$$M^{**} = M,$$

by 2.64 and 2.25.

2.66 *Let* $N$ *be a chain-group on* $E$ *over a commutative ring* $R$ *with a unit element and no divisors of zero. Then*

$$(M(N))^* = M(N^*),$$

by 2.43, 2.64 and 2.25. In particular we have

2.67 *For any finite graph* $G$,

$$(P(G))^* = B(G),$$

by 2.52.

## 3. Structure of Matroids

### 3.1. Subgraphs and Contractions

Let $G$ be a graph and let $S$ be a subset of $E(G)$. There are two specially important subgraphs of $G$ associated with $S$. One is the *spanning subgraph* $G:S$, for which $V(G:S) = V(G)$ and $E(G:S) = S$. The other is the *reduction* of $G$ to $S$, defined by the edges of $G$ and their incident vertices. We denote it by $G \cdot S$.

We note that $G \cdot S$ has no isolated vertices. For this reason we shall usuall refer the reductions of $G$ to the other subgraphs. Fo the addition or removal of an isolated vertex in a graph $G$ does not affect the 1-cycles or coboundaries.

The *contraction* $G$ ctr $S$ of $G$ to $S$ is a graph whose vertices are the components of $G:(E(G) - S)$ and whose edges are the members of $S$. The ends of an edge in $G$ ctr $S$ are the two components, possibly coincident, of $G:(E(G) - S)$ which include the ends of the edge in $G$.

If the identities of the vertices of $G$ are not important we may say, less precisely, that $G$ ctr $S$ is formed from $G$ by contracting each component of $G:(E(G)-S)$ to a single vertex. Alternatively we may say that $G$ ctr $S$ is obtained from $G$ by a sequence of "elementary contractions" in each of which one edge of $G$ is contracted to a single vertex.

By deleting the isolated vertices, if any, of $G$ ctr $S$ we obtain from it the *reduced contraction* $G\times S=(G \text{ ctr } S)\cdot S$ of $G$ to $S$.

We go on to derive some corresponding operations on chain-groups and matroids.

### 3.2. Minors of Chain-Groups

Let $N$ be a chain-group on a set $E$ over a commutative ring $R$ having a unit element and no divisors of zero.

Suppose $S\subseteq E$. If $f$ is any chain of $N$ we define its *restriction* to $S$ as the chain $g$ on $S$ such that $g(a)=f(a)$ for each $a\epsilon S$.

The restrictions to $S$ of the chains of $N$ constitute a chain-group on $S$ over $R$. We call this the *reduction* of $N$ to $S$ and denote it by $N\cdot S$. Another chain-group on $S$ over $R$ is given by the restrictions to $S$ of those chains $f$ of $N$ for which $\|f\|\subseteq S$. This is the *contraction* of $N$ to $S$, denoted by $N\times S$.

3.21 $N\times S\subseteq N\cdot S$.

The proofs of the two following theorems are simple exercises in graph theory.

3.22 *Let $N$ be the group of coboundaries over $R$ of a graph $G$. Then $N\cdot S$ is the group of coboundaries over $R$ of $G\cdot S$.*

3.23 *Let $N$ be as in 3.22. Then $N\times S$ is the group of coboundaries over $R$ of $G\times S$.*

Returning to the general case we suppose $T\subseteq S\subseteq E$. We can now establish the following identities

3.241 $$(N\times S)\times T=N\times T,$$

3.242 $$(N\cdot S)\cdot T=N\cdot T,$$

3.243 $$(N\cdot S)\times T=(N\times(E-(S-T)))\cdot T,$$

3.244 $$(N\times S)\cdot T=(N\cdot(E-(S-T)))\times T.$$

The first two of these follow immediately from the definitions. To prove 3.243 we observe that each side of the formula represents the restrictions to $T$ of those chains of $N$ which have only zero coefficients in $S-T$. We can obtain 3.244 by writing $E-(S-T)$ for $S$ in 3.243.

We refer to a chain-group of the form $(N\cdot S)\times T$ as a *minor* of $N$. The minors of $N$ include $N$ itself and all its reductions and contractions. For $N=(N\cdot E)\times E$, $N\cdot S=(N\cdot S)\times S$, and $N\times S=(N\cdot E)\times S$. From the identities 3.241 we deduce

3.25 *Every minor of a minor of $N$ is a minor of $N$.*

3.26 *If $N$ is regular then every minor of $N$ is regular.*

PROOF: Suppose $S\subseteq E$. Let $f$ be an elementary chain of $N\times S$. Then $f$ is the restriction to $S$ of an elementary chain $f_1$ of $N$ having the same domain as $f$. But $f_1$ is an integral multiple of a primitive chain of $N$, since $N$ is regular. This implies that $f$ is an integral multiple of a primitive chain of $N\times S$. We deduce that $N\times S$ is regular.

Now let $h$ be an elementary chain of $N\cdot S$. There is a chain $h_1$ of $N$ having $h$ as its restriction to $S$. Choose $h_1$ so that $\|h_1\|$ has the least possible number of cells. There is an elementary chain $h_2$ of $N$ such that $\|h_2\|\subseteq\|h_1\|$. By the choice of $h_1$ we may suppose $\|h_2\|\cap S$ is non-null. Since $h_2$ is a multiple of a primitive chain of $N$ we deduce that there is a nonzero chain $k$ of $N\cdot S$, with coefficients restricted to the values 1, $-1$ and 0, such that $\|k\|\subseteq\|h\|$. But then $\|k\|=\|h\|$ since $h$ is elementary, and therefore $k$ is a primitive chain of $N\cdot S$. Since $h$ is elementary it must be a multiple of $k$. We deduce that $N\cdot S$ is regular.

The theorem follows from these two results.

Minors of dual chain-groups are related by the following identity.

3.27 $$(N\cdot S)^*=N^*\times S.$$

PROOF: Let $f$ be a chain on $S$. It belongs to $(N\cdot S)^*$ if and only if it is orthogonal to every chain of $N\cdot S$, that is if and only if the chain $g$ on $E$ which satisfies $\|g\|=\|f\|$ and has $f$ as its restriction to $S$ is orthogonal to every member of $N$. But this condition holds if and only if $g\epsilon N^*$, that is $f\epsilon N^*\times S$.

In the cases of chief interest to us we can supplement this result as follows.

3.28 *Suppose that $R$ is a field or that $N$ is regular. Then*

$$(N\times S)^*=N^*\cdot S,$$

*for each $S\subseteq E$.*

PROOF: Writing $N^*$ for $N$ in 3.27, and then taking dual chain-groups we obtain

$$(N^{**}\times S)^*=(N^*\cdot S)^{**}.$$

But under the conditions of the theorem $N^{**}=N$, by 2.46. Similarly $(N^*\cdot S)^{**}=N^*\cdot S$, by 2.47 and 3.26. The theorem follows

3.29 *Under the conditions of 3.28 the duals of the minors of $N$ are the minors of $N^*$* (by 3.27 and 3.28).

### 3.3. Minors of Matroids

This section is closely analogous to 3.2.

Let $M$ be a matroid on a set $E$, and suppose $S\subseteq E$.

Let $L$ be the class of non-null intersections with $S$ of atoms of $M$, and let $M\cdot S$ be the class of minimal members of $L$. Since $M$ satisfies Axiom II it is clear that $L$ does so too. Hence $M\cdot S$ is a matroid on $S$, by 1.12. We call it the *reduction* of $M$ to $S$.

Let $M\times S$ be the class of all atoms $X$ of $M$ such that $X\subseteq S$. Then $M\times S$ satisfies Axioms I and II. It is thus a matroid on $S$. We call it the *contraction* of $M$ to $S$.

By a comparison of definitions we have the following theorem

3.31 *Let $N$ be a chain-group on $E$ over a ring $R$ with a unit element and no divisors of zero. Then*

$$M(N\times S)=M(N)\times S,$$

$$M(N\cdot S)=M(N)\cdot S.$$

Now take $N$ to be the group of coboundaries over $R$ of a graph $G$. Then by 3.22, 3.23 and 3.31 we have

3.321 $$B(G\cdot S)=B(G)\cdot S,$$

3.322 $$B(G\times S)=B(G)\times S.$$

Here $B(G)$ is the bond-matroid of $G$ defined in section 2.5.

Returning to the general case we suppose $T\subseteq S\subseteq E$. We establish the following identities

3.331 $$(M\times S)\times T=M\times T,$$

3.332 $$(M\cdot S)\cdot T=M\cdot T,$$

3.333 $$(M\cdot S)\times T=(M\times(E-(S-T)))\cdot T,$$

3.334 $$(M\times S)\cdot T=(M\cdot(E-(S-T)))\times T.$$

The first of these follows at once from the definitions.

To prove 3.332 we observe that if $X\epsilon(M\cdot S)\cdot T$ then $X$ is a non-null intersection with $T$ of an atom of $M\cdot S$, and therefore a non-null intersection with $T$ of an atom of $M$. Hence there exists $Y\epsilon M\cdot T$ such that $Y\subseteq X$. Conversely suppose $Y\epsilon M\cdot T$. Then $Y$ is a non-null intersection with $T$ of an atom $Z$ of $M$. But $Z$ contains an atom $Z_1$ of $M\cdot S$ meeting $T$, by 1.11. Similarly $Z_1$ contains an atom $X$ of $(M\cdot S)\cdot T$, and we have $X\subseteq Y$. Applying Axiom I to these results we have 3.332.

We prove 3.333 in a similar way. Suppose $X\epsilon(M\cdot S)\times T$. Then $X$ is an atom of $M\cdot S$ contained in $T$. Accordingly $M$ has an atom $X_1$ such that $X_1\cap T=X$ and $X_1\cap(S-T)$ is null. But then $X_1$ is an atom of $M\times(E-(S-T))$ and hence there is an atom $Y$ of $(M\times(E-(S-T)))\cdot T$ such that $Y\subseteq X$. Conversely suppose $Y\epsilon(M\times(E-(S-T)))\cdot T$. Then $Y$ is the intersection with $T$ of an atom $Y_1$ of $M\times(E-(S-T))$, and $Y_1$ is an atom of $M$ such that $Y_1\cap(S-T)$ is null. Hence there is an atom $X$ of $M\cdot S$ which is contained in $Y_1$. We have in fact $X\subseteq Y\subseteq T$. Hence $(M\cdot S)\times T$ has an atom $X$ such that $X\subseteq Y$. Applying Axiom I to these results we establish 3.333.

We obtain 3.334 by writing $E-(S-T)$ for $S$ in 3.333.

We refer to a matroid of the form $(M\cdot S)\times T$ as a *minor* of $M$. The minors of $M$ evidently include $M$ itself and all its reductions and contractions. From the four identities just proved we deduce

3.34 *Every minor of a minor of $M$ is a minor of $M$.*

Minors of dual matroids are related by the following identities

3.351 $$(M\cdot S)^*=M^*\times S,$$

3.352 $$(M\times S)^*=M^*\cdot S.$$

To prove 3.351 suppose $X\epsilon(M\cdot S)^*$. Then $X\subseteq S$ and $X$ is orthogonal to every atom of $M\cdot S$. Now the class of intersections with $S$ of atoms of $M$ satisfies Axiom II and its minimal members constitute $M\cdot S$. If $X$ meets one of these intersections in a single cell, then it meets some atom of $M\cdot S$ in a single cell, by 1.11. Hence $X$ is orthogonal to every atom of $M$. Accordingly there exists $Y\epsilon M^*$ such that $Y\subseteq X$. But then $Y\epsilon M^*\times S$.

Conversely suppose $Y\epsilon M^*\times S$. Then $Y\subseteq S$ and $Y\epsilon M^*$. Hence $Y$ is orthogonal to every member of $M$ and therefore to every member of $M\cdot S$. Accordingly there exists $X\epsilon(M\cdot S)^*$ such that $X\subseteq Y$.

Applying Axiom I to these results we obtain 3.351.

Writing $M^*$ for $M$ in 3.351 and taking dual matroids we obtain

$$(M^*\cdot S)^{**}=(M^{**}\times S)^*$$

By 2.65 this reduces to 3.352.

3.36 *The minors of $M^*$ are the duals of the minors of $M$.*

This follows from 3.351 and 3.352, with the help of 3.334.

The foregoing results can be applied to the bond-matroid and polygon-matroid of a graph $G$, these being duals by 2.67. On taking the dual forms of 3.321 and 3.322 we obtain the following identities.

3.371 $$P(G\cdot S)=P(G)\times S,$$

3.372 $$P(G\times S)=P(G)\cdot S.$$

## 3.4. Connection in Matroids

Consider a matroid $M$ on a set $E$. We define a *separator* of $M$ as a subset $S$ of $E$ such that each atom of $M$ is contained either in $S$ or in $E-S$. Evidently any union or intersection of separators of $M$ is a separator of $M$, and the complement of a separator of $M$ is also a separator.

We refer to the minimal non-null separators of $M$ as its *elementary* separators. From the foregoing observations we deduce

3.41 *The elementary separators of $M$ are disjoint non-null subsets of $E$ whose union is $E$.*

3.42 *Suppose $S\subseteq E$. Then $S$ is a separator of $M$ if and only if*

$$M\cdot S=M\times S.$$

PROOF: Let $S$ be a separator of $M$. Then an atom of $M$ has a non-null intersection with $S$ if and only if it is itself a subset of $S$. It follows that $M\times S$ and $M\cdot S$ are identical.

Conversely suppose $M\cdot S=M\times S$, and let $Y$ be any atom of $M$. If it meets $S$ it contains an atom of $M\cdot S$, that is $M\times S$. It is then itself an atom of $M\times S$ and $M$, by Axiom I. Thus $Y\subseteq S$. We deduce that $S$ is a separator of $M$.

3.43 *Let $S$ be a separator of $M$. Then, for each $T\subseteq E$, $S\cap T$ is a separator of both $M\cdot T$ and $M\times T$.*

PROOF: Let $Y$ be an atom of $M \cdot T$ or $M \times T$. There is an atom $Z$ of $M$ such that $Y = Z \cap T$. Either $Z \subseteq S$ or $Z \subseteq E - S$. Hence $Y$ is contained either in $S \cap T$ or in $(E-S) \cap T$. The theorem follows.

3.44 *Let $S$ be a separator of $M$, and let $T$ be a separator of $M \cdot S$, that is $M \times S$ by 3.42. Then $T$ is a separator of $M$.*

PROOF: $M \times T = (M \times S) \times T = (M \times S) \cdot T = (M \cdot S) \cdot T = M \cdot T$, by 3.331, 3.332 and 3.42. Hence $T$ is a separator of $M$, by 3.42.

3.45 *The separators of $M^*$ are the separators of $M$.*

PROOF: If $M \cdot S = M \times S$ we have $M^* \times S = M^* \cdot S$, by 3.351 and 3.352. Similarly if $M^* \cdot S = M^* \times S$ we have $M \times S = M \cdot S$, by 3.351, 3.352, and 2.65. The theorem follows, by 3.42.

If $S$ is an elementary separator of $M$ we refer to the matroid $M \cdot S$, that is $M \times S$, as a *component* of $M$.

The matroid $M$ is said to be *connected* if it has no separators other than the trivial ones, $E$ and its null subset. Thus $M$ is connected if and only if either $E$ is null or $M$ has just one elementary separator.

3.46 *If a minor $(M \cdot S) \times T$ of $M$ is connected, and $T$ is not null, then $T$ is a subset of an elementary separator of $M$.*

We obtain this result by two applications of 3.43.

3.47 *Any component of $M$ is connected (by 3.44).*

3.48 *Let $M \times S$ and $M \times T$ be connected reductions of $M$ such that $S \cap T$ is non-null. Then $M \times (S \cup T)$ is connected.*

PROOF: There is an elementary separator $Z$ of $M \times (S \cup T)$ which meets $S \cap T$, by 3.41. It contains both $S$ and $T$, by 3.331 and 3.43. Hence $Z = S \cup T$, and the theorem follows.

3.49 *The components of $M^*$ are the duals of the components of $M$.*

This follows from 3.45, with the help of 3.351, 3.352 and 3.42.

Consider a graph $G$ without isolated vertices. It is called *nonseparable* or *cyclically connected* if it has the following property: for any two complementary non-null subsets $S$ and $T$ of $E(G)$ there exists a polygon of $G$ whose edge-set meets both $S$ and $T$.

The property of cyclic connection in graphs corresponds to that of connection in matroids. For it is clear that $G$ is nonseparable if and only if $P(G)$ is connected. Hence $G$ is nonseparable if and only if $B(G)$ is connected, by 2.67 and 3.45.

There is no property of matroids in general which corresponds in this way to ordinary connection for graphs.

### 3.5. Properties of Rank

Let $M$ be a matroid on a set $E$. From the definition of rank given in section 2.2 we have the following theorem.

3.51 $0 \leq r(M) \leq |E|$. *Moreover $r(M) = 0$ if and only if $M$ has no atom, and $r(M) = |E|$ if and only if each cell of $E$ constitutes by itself an atom of $M$.*

From 2.64 we deduce

$$3.52 \qquad r(M) + r(M^*) = |E|.$$

We now supplement these results with some formulae involving the ranks of minors of $M$.

3.53 *Let $S$ be any subset of $E$. Let $C$ be a dendroid of $M \times S$ and let $D$ be a dendroid of $M \cdot (E-S)$. Then $C \cup D$ is a dendroid of $M$.*

PROOF: Let $a$ be any cell of $C \cup D$.

If $a \in C$ we define $F_a$ as the atom $J(C, a)$ of $M \times S$. It is of course also an atom of $M$, and $|F_a \cap D| = 0$.

If $a \in D$ we define $F_a$ as an atom of $M$ for which $F_a \cap D = \{a\}$ and for which $|F_a \cap C|$ has the least value consistent with this condition. Such a set exists since the atom $J(D, a)$ of $M \cdot (E-S)$ is the intersection with $E-S$ of some atom of $M$. We show that in fact $|F_a \cap C| = 0$. For suppose $b \in F_a \cap C$. Applying Axiom II to $F_a$ and the atom $Y = J(C, b)$ of $M \times S$ we find that there is an atom $F$ of $M$ such that $a \in F \subseteq (F_a \cup Y) - \{b\}$. But then $F \cap D = \{a\}$ and $|F \cap C| < |F_a \cap C|$, which is contrary to the definition of $F_a$.

We conclude that for each $a \in C \cup D$ there is an atom $F_a$ of $M$ such that $F_a \cap (C \cup D) = \{a\}$.

On the other hand any atom of $M$ meets $D$ if it has a non-null intersection with $E-S$, and meets $C$ if it is contained in $S$. Hence $C \cup D$ is a dendroid of $M$.

We note the following corollaries.

$$3.54 \qquad r(M \times S) + r(M \cdot (E-S)) = r(M).$$

3.55 *Suppose $a \in E$. Then $r(M) - r(M \times (E - \{a\})) = 1$ or 0 according as $a$ does or does not belong to some atom of $M$.*

The next theorem finds frequent applications in the next chapter.

3.56 *Let $S$ and $T$ be subsets of $E$. Then*

$$r(M \times (S \cup T)) + r(M \times (S \cap T)) \geq r(M \times S) + r(M \times T).$$

PROOF: Let $C$ be a dendroid of $M \times (S \cap T)$. Let $D_1$ be a dendroid of $(M \times S) \cdot (S-T)$ and $D_2$ a dendroid of $(M \times T) \cdot (T-S)$. Write $U = C \cup D_1 \cup D_2$.

Now $C \cup D_1$ is a dendroid of $M \times S$, by 3.331 and 3.53. Hence if $a \in C \cup D_1$ there is an atom $F_a$ of $M \times (S \cup T)$, contained in $S$, which meets $U$ only in the cell $a$. An analogous result holds for $a \in C \cup D_2$. We deduce that each cell of $(M \times (S \cup T)) \cdot U$ constitutes an atom of that matroid. Accordingly

$$r((M \times (S \cup T)) \cdot U) = |C \cup D_1 \cup D_2|,$$
$$= r(M \times S) + r(M \times T) - r(M \times (S \cap T)).$$

The theorem follows, by 3.54.

3.57 *$r(M) = 1$ if and only if $M$ has just one atom.*

PROOF: If $M$ has just one atom $X$, then any cell of $X$ constitutes a dendroid of $M$, and $r(M) = 1$. Conversely if $r(M) = 1$ then $M$ has a dendroid $D$ consisting of a single cell $a$. But $M$ has only one atom $J(D, a)$ which includes $a$.

## 4. Geometry of Matroids

### 4.1. Flats

Let $M$ be a matroid on a set $E$. We propose to study the system of contractions $M \times S$ of $M$. We abbreviate the expression $r(M \times S)$ as $rS$. For the sake of geometrical analogies to be explained below we define the *dimension* $dS$ of $S$ as the number $rS - 1$.

A subset $S$ of $E$ is called a *flat* of $M$ if it is a union of atoms of $M$. The null subset of $E$ is counted as a null union of atoms, and therefore a flat. For any subset $S$ of $E$ there is an associated flat $\langle S \rangle$, defined as the union of all the atoms of $M$ contained in $S$, that is the union of the atoms of $M \times S$. By 3.54 we have

4.11 $$d\langle Z \rangle = dZ = rZ - 1.$$

In what follows we are concerned with the lattice of flats of $M$, as partially ordered by the inclusion relation.

4.12 *The only flat of $M$ with negative dimension is the null flat $\phi$, for which $d\phi = -1$. The flats of dimension 0 are the atoms of $M$.*

This follows from 3.51 and 3.57. It is convenient to take over some more geometrical terminology at this stage. We say that the flat $S$ is *on* the flat $T$ if either $S \subseteq T$ or $T \subseteq S$. The atoms of $M$ are called also its *points*, and flats of dimensions 1 and 2 are *lines* and *planes* of $M$ respectively. More generally a flat of dimension $k$ is called a $k$-flat.

4.13 *If $S$ is a flat of $M$ and $a \in S$, then*

$$d\langle S - \{a\} \rangle = dS - 1,$$

by 3.55.

4.14 *If $S$ and $T$ are flats of $M$ such that $S \subset T$, then $dS < dT$. Moreover there is a flat $U$ of $M$ such that $S \subset U \subseteq T$ and $dU = dS + 1$.*

PROOF: The first statement follows from 3.54. To prove the second we define $U$ as a flat of $M$ satisfying $S \subset U \subseteq T$ and such that $|U|$ has the least value consistent with this condition. Choose $a \in U - S$. Then $S \subseteq \langle U - \{a\} \rangle \subset T$, and therefore $S = \langle U - \{a\} \rangle$ by the choice of $U$. Hence $dU = dS + 1$, by 4.13.

4.15 *Let $S$ and $T$ be flats of $M$ such that $S \subseteq T$. Then there exists a flat $U$ of $M$ such that $U \subseteq T$, $\langle U \cap S \rangle = \phi$ and $dU = dT - dS - 1$.*

PROOF: Write $S_0 = S$ and $T_0 = T$. If possible choose $a_0 \in S_0$ and write $S_1 = \langle S_0 - \{a_0\} \rangle$, $T_1 = \langle T_0 - \{a_0\} \rangle$. It is clear that $S_1 \subseteq T_1$. If possible choose $a_1 \in S_1$ and write $S_2 = \langle S_1 - \{a_1\} \rangle$, $T_2 = \langle T_1 - \{a_1\} \rangle$. Then $S_2 \subseteq T_2$.

Continue this process until it terminates. By 4.13 this will be with $S_k$ and $T_k$, where $k = dS + 1$ and $S_k = \phi$. Applying 4.13 to the sequence of the $T_i$ we find that $dT_k = dT - dS - 1$. But $\langle T_k \cap S \rangle = \langle S - \{a_0, a_1, \ldots, a_{k-1}\} \rangle = S_k = \phi$. Hence the theorem is satisfied with $U = T_k$.

4.16 *Let $S$ and $T$ be flats of $M$. Then*

$$d(S \cup T) + d\langle S \cap T \rangle \geq dS + dT,$$

by 3.56.

Many "geometrical" results can be deduced from 4.14 and 4.16. An example follows.

4.171 *Let $L_1$ and $L_2$ be distinct lines of $M$ on the same plane $P$. Then $L_1$ and $L_2$ have just one common point.*

PROOF: We have $L_1 \subseteq L_1 \cup L_2 \subseteq P$. Hence $L_1 \cup L_2 = P$, by 4.14. Another application of 4.14 shows that $d\langle L_1 \cap L_2 \rangle \leq 0$ since $L_1 \cap L_2 \subset L_1$. But $d\langle L_1 \cap L_2 \rangle \geq 0$ by 4.16. Hence $d\langle L_1 \cap L_2 \rangle = 0$, and the theorem follows by 4.12.

The two following theorems can be proved in a similar way.

4.172 *Let $P_1$ and $P_2$ be distinct planes on the same 3-flat $F$ of $M$. Then $\langle P_1 \cap P_2 \rangle$ is a line on $F$.*

4.173 *Let $P$ be a plane and $L$ a line on the same 3-flat $F$. Suppose $L$ is not on $P$. Then $\langle P \cap L \rangle$ is a point on $E$.*

We observe that if $G$ is any graph, then the flats of $P(G)$ correspond to those subgraphs of $G$ which are unions of polygons.

### 4.2. Connected Flats

A flat $S$ of $M$ is called "connected" if $M \times S$ is a connected matroid. We then refer to the separators of $M \times S$ as the "separators of $S$". The lattice of connected flats of $M$ has some very interesting properties. We begin this section with a study of the line.

4.21 *Let $L$ be a line of $M$, and suppose $a \in L$. Then $\langle L - \{a\} \rangle$ is the only point on $L$ which does not include $a$.*

PROOF: $\langle L - \{a\} \rangle$ is a point on $L$, by 4.13. But if $X$ is any point on $L$ not including $a$ we have $X \subseteq \langle L - \{a\} \rangle$ and therefore $X = \langle L - \{a\} \rangle$ by Axiom I.

4.22 *Any line $L$ of $M$ is on two distinct points. If $X$ and $Y$ are distinct points on $L$ then $L = X \cup Y$. Moreover $X \cap Y$ is non-null if and only if $L$ is connected.*

PROOF: Choose $a \in L$ and $b \in \langle L - \{a\} \rangle$. Then $\langle L - \{a\} \rangle$ and $\langle L - \{b\} \rangle$ are distinct points of $L$.

Let $X$ and $Y$ be distinct points on $L$. Then $X \subset X \cup Y \subseteq L$. Hence $X \cup Y = L$, by 4.14. If $X \cap Y$ is non-null then $L$ is connected, by 3.48. If $X \cap Y$ is null then either $X$ and $Y$ are nontrivial separators of $L$, or there is a point $Z$ on $L$ meeting both $X$ and $Y$. In the latter case $X \subset X \cup Z \subset X \cup Y = L$, by Axiom I. But this is impossible, by 4.14.

4.23 *A disconnected line is on just two points, and a connected line is on at least three points.*

PROOF: By 4.22 any two distinct points on a disconnected line $L$ are disjoint, and have $L$ as their union. Hence $L$ has at most two points, and therefore just two, by 4.22.

By 4.22 any connected line $L$ has two distinct points $X$ and $Y$, and we can find $a \in X \cap Y$. By 4.21 $\langle L - \{a\} \rangle$ is a point on $L$ which is distinct from $X$ and $Y$.

It should perhaps be pointed out that two distinct points $X$ and $Y$ of a matroid $M$ are not necessarily on a common line. There is indeed a flat $X \cup Y$, but its dimension may exceed 1. However the following theorem is sometimes helpful.

4.24 *Let $T$ be a separator of a flat $S$ of $M$. Let $X$ and $Y$ be points on $S$ such that $X \subseteq T$ and $Y \subseteq S - T$. Then $X \cup Y$ is a disconnected line of $M$.*

PROOF: $X$ and $Y$ are separators of the flat $X \cup Y$, by 3.43. Hence $d(X \cup Y) = dX + dY - 1 = 1$, by 3.42 and 3.54. Thus $X \cup Y$ is a disconnected line.

We go on to prove two important theorems, 4.26 and 4.27, about connected flats in general. We need the auxiliary result

4.25 *Let $S$ and $T$ be connected flats of $M$ such that $S \subset T$. Then there exists a connected $(dS+1)$-flat $U$ of $M$ such that $S \subset U \subseteq T$.*

PROOF: Since $T$ is connected we can find a point $X$ of $M$ such that $X \subseteq T$ and $X$ meets both $S$ and $T-S$. Choose such an $X$ so that $|S \cup X|$ has the least possible value.

$S \cup X$ is a connected flat of $M$, by 3.48. Its dimension exceeds $dS$, by 4.14.

Suppose $d(S \cup X) > dS + 1$. Choose $a \in X - S$. Then

$$d\,\langle (S \cup X) - \{a\} \rangle \geqslant dS + 1,$$

by 4.13. Hence there is a point $Y$ of $M$ such that $Y \subseteq (S \cup X) - \{a\}$ and $Y$ meets $X - S$. But $Y \cap S = \phi$ by the choice of $X$. Hence $Y \subset X$, which is contrary to Axiom I. We deduce that in fact $d(S \cup X) = dS + 1$. Hence the theorem holds with $U = S \cup X$.

4.26 *Let $S$ be a connected $d$-flat on a connected $(d+2)$-flat $T$ of $M$. Then there exist distinct connected $(d+1)$-flats $U$ and $V$ of $M$ such that $S = \langle U \cap V \rangle$ and $T = U \cup V$.*

PROOF: The following argument is illustrated by figure 4A for the case $d = 0$, in which $S$ is a point and $T$ is a connected plane.

By 4.25 there is a connected $(d+1)$-flat $U$ which is on both $S$ and $T$. Choose $a \in U - S$ and write $W = \langle T - \{a\} \rangle$. By 4.13 $W$ is another $(d+1)$-flat on both $S$ and $T$. By 4.15 there is a line $L$ on $T$ having no common point with $S$. It meets $U$ and $W$ in points $X$ and $Z$, respectively, by 4.16. By 4.14 we have $S \cup X = U$ and $S \cup Z = W$. Hence $Z$ is not on $U$ and $U \cup Z = T$, by further applications of 4.14.

Assume $W$ is not connected. Then $S \cap Z = \phi$, by 3.48.

Suppose $U \cap Z = \phi$. By the connection of $T$, that is $U \cup Z$, there is an atom $Z'$ of $M$ such that $Z' \subseteq T$ and $Z'$ meets both $U$ and $Z$. Then $U \subset U \cup Z' \subset U \cup Z = T$, by Axiom I. But this is impossible by 4.14. We deduce that $U \cap Z$ is non-null. A similar argument in which $X$, $S$, and $U$ replace $Z$, $U$, and $T$ respectively shows that $X \cap S$ is not null. Choose $b \in Z \cap U$ and $c \in X \cap S$.

Write $V = \langle T - \{b\} \rangle$. By 4.13 $V$ is a $(d+1)$-flat. It is on $S$, since $b \in Z$ and $S \cap Z = \phi$. By 4.16 it has a common point $Y$ with $L$. Since $L$ and $S$ have no common point we deduce from 4.14 that $V = S \cup Y$. Moreover $V$ is distinct from $U$ and $W$, since $b \in Z \cap U \subseteq W \cap U$.

Now $c \in S \cap X \subseteq S \cap L = S \cap (Y \cup Z)$, by 4.22, $= S \cap Y$, by our assumption. Hence $V$ is connected, by 3.48.

In the remaining case, in which $W$ is connected, we write $V = W$.

In either case we have two distinct connected $(d+1)$-flats $U$ and $V$ of $M$, each of which is on both $S$ and $T$. Hence

$$S \subseteq \langle U \cap V \rangle \subset U \subset U \cup V \subseteq T,$$

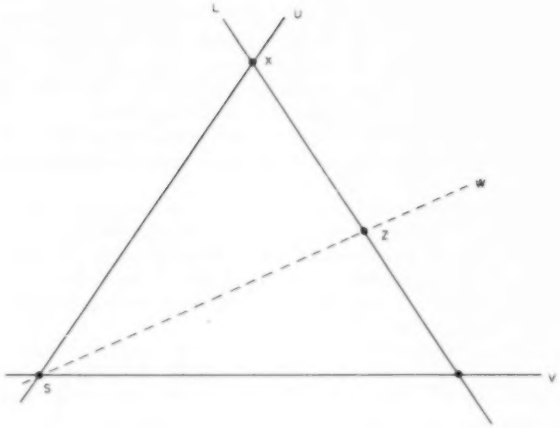



FIGURE 4A

since neither of $U$ or $V$ contains the other, by 4.14. Another application of 4.14 shows that $S = \langle U \cap V \rangle$ and $T = U \cup V$.

4.27 *Let $S$, $T$, and $U$ be flats of $M$ such that $S$ and $T$ are connected, $S \cup U \subseteq T$ and $\langle S \cap U \rangle = \phi$. Then there exists a connected flat $R$ of $M$ such that $S \subseteq R \subseteq T$, $\langle R \cap U \rangle = \phi$, and $dR = dT - dU - 1$.*

PROOF: If possible choose $S$, $T$, and $U$ so that the theorem fails and $dU$ has the least value consistent with this condition. Then $dU > -1$, since otherwise the theorem holds with $T = R$.

Let $W$ be a connected flat of $M$ of greatest possible dimension such that $S \subseteq W \subset T$ and $W$ does not contain $U$. Suppose $dW < dT - 1$. Then, by 4.26, there are distinct connected $(dW+1)$-flats $K$ and $L$ on $T$ such that $\langle K \cap L \rangle = W$. But then $K$ and $L$ cannot both contain $U$, and this is contrary to the choice of $W$. We deduce, using 4.14, that $dW = dT - 1$. We note that $d\,\langle U \cap W \rangle < dU$, by 4.14.

By the choice of $S$, $T$, and $U$ there is a connected flat $R$ of $M$ such that $S \subseteq R \subseteq W \subset T$, $\langle R \cap \langle U \cap W \rangle \rangle = \phi$ and $dR = dW - d\langle U \cap W \rangle - 1$. But then $\langle R \cap U \rangle = \phi$ and $dR \geqslant dT - dU - 1$. However $dR \leqslant dT - dU - 1$, by 4.14 and 4.16. We deduce that $dR = dT - dU - 1$. Thus the theorem holds for $S$, $T$, and $U$, which is a contradiction.

It follows that the theorem is true in general.

We shall also need the following more special theorem.

4.28 *Let $L$ be a disconnected line on a connected $d$-flat $S$ of $M$, where $dS > 1$. Then there exists a connected plane $P$ of $M$ such that $L \subset P \subseteq S$.*

PROOF: Let the two points on $L$ be $X$ and $Y$. (See 4.23.) Let $P$ be a connected flat of $M$, of least possible dimension, such that $L \subset P \subseteq S$. Assume $dP > 2$.

Suppose first that there is a second disconnected line $L'$ on $X$ and $P$. Let its point other than $X$ be $Z$. Then $X$, $Y$, and $Z$ are distinct points, by 4.14. By 4.27 there is a connected $(dP-2)$-flat $U$ on $Y$ and $P$ which has no point in common with $L'$. By 4.26

there are distinct connected $(dP-1)$-flats $V$ and $W$ on $P$ such that $\langle V \cap W \rangle = U$. These two flats meet $L'$ in distinct points, by 4.14. Since there are only two points on $L'$ we may suppose $X$ is on $V$. But then $L$ is on $V$ and the definition of $P$ is contradicted.

We deduce that there is no second disconnected line on $X$ and $P$, and similarly no second disconnected line on $Y$ and $P$.

Choose $a \epsilon P - L$ and write $R = \langle P - \{a\} \rangle$. Then $L \subseteq R$, and $dR = dP - 1$ by 4.13. By the definition of $P$ the flat $R$ is disconnected. But the only possible non-trivial separators of $R$ are $X$ and $Y$, by 4.24. Accordingly $R = L$ and $dP = 2$, contrary to assumption. The theorem follows.

4.281 *Let $L$ be a disconnected line on a plane $P$ of $M$. Let $X$ and $Y$ be its two points, and let $Z$ be any other point on $P$. Then $X \cup Z$ and $Y \cup Z$ are connected lines, the only lines of $M$ which are on both $Z$ and $P$.*

PROOF: Any line on $P$ contains $X$ or $Y$, by 4.171. Hence, by 4.22, the only flats on $Z$ and $P$ which can be lines are $X \cup Z$ and $Y \cup Z$. They must in fact be connected lines, by 4.26.

4.282 *Let $L$ be a disconnected line on a connected plane $P$ of $M$. Then every line on $P$ other than $L$ is connected.*

PROOF: Let $L'$ be such a line. It is on some point $Z$ not contained in $L$, by 4.22. Hence it is connected, by 4.281.

The foregoing results can be applied to the polygon-matroid of a graph $G$ to obtain some rather simple results in graph theory. A set $S$ is a connected flat of $P(G)$ if and only if $G \cdot S$, whose polygon matroid is $P(G) \times S$ by 3.371, is nonseparable. The rank of $P(G)$ is the *cyclomatic number* of $G$, that is the least number $p_1(G)$ of edges of $G$ which must be deleted in order to destroy every polygon.

Two connected flats $S$ and $T$ of $P(G)$ such that $\langle S \cap T \rangle = \phi$ correspond to two nonseparable subgraphs of $G$ which have no common polygon.

## 4.3. Linear Subclasses

Let $M$ be a matroid on a set $E$, and let $C$ be any class of points of $M$. We call $C$ a *linear subclass* of $M$ if it has the following property. If two distinct points $X$ and $Y$ of $C$ are on a common line $L$ in $M$, then every point on $L$ belongs to $C$. In earlier work on matroids a linear subclass is called a "convex subclass".

An obvious example of a linear subclass of $M$ is provided by the class of all points on a given flat.

Another important example arises as follows. Let us say that a subset $S$ of $E$ *cuts* another subset $T$ of $E$ if both $S \cap T$ and $T - S$ are non-null. Then we have the following theorem.

4.31 *Let $S$ be any subset of $E$. Let $C$ be the class of all atoms of $M$ which do not cut $S$. Then $C$ is a linear subclass of $M$.*

PROOF: Let $X$ and $Y$ be points of $C$ on a common line $L$. Let $Z$ be any point on $L$. We have $Z \subset X \cup Y = L$, by 4.22.

If $X \cap S$ and $Y \cap S$ are null we have $Z \cap S = \phi$ and therefore $Z \epsilon C$. If $S \subseteq X$ and $S \subseteq Y$ we have $S \subseteq Z$ and therefore $Z \epsilon C$.

In the remaining case we may suppose $X \cap S = \phi$ and $S \subseteq Y$. Hence $L$ is disconnected, by 4.22. Accordingly $Z = X$ or $Y$, by 4.23. Again we have $Z \epsilon C$.

The theorem follows.

We have so far presented the theory of matroids as something more general than that of graphs, having its own theorems from which the results of graph theory can be derived as special cases. It is somewhat surprising therefore to find that we must now use the technique of graph theory in order to study linear subclasses. However each matroid does have an associated graph. The vertices are the points of the matroid, and two points are said to be *adjacent* if and only if they are distinct points of the same connected line.

A *path* in $M$ is a finite sequence

$$P = (X_1, \ldots, X_k)$$

of one or more points of $M$, not necessarily all distinct, such that any two consecutive terms are adjacent (and therefore distinct) points of $M$. We refer to $X_1$ and $X_k$ as the *origin* and *terminus* of $P$ respectively. If they coincide we call $P$ *re-entrant*. If $P$ has only one term it is *degenerate*. The *length* $s(P)$ of $P$ is one less that the number of terms of $P$.

If $P = (X_1, \ldots, X_k)$ and $Q = (X_k, \ldots, X_m)$ are paths of $M$ such that the origin of $Q$ is the terminus of $P$, then we define their *product* $PQ$ as the path $(X_1, \ldots, X_k, \ldots, X_m)$. Multiplication of paths is clearly associative. We may therefore write a product $(PQ)R$ or $P(QR)$ simply as $PQR$

If $P$ is any path of $M$ we write $P^{-1}$ for the path obtained by taking the terms of $P$ in reverse order. The following rules are obvious.

4.321 $$(P^{-1})^{-1} = P,$$

4.322 $$(PQ)^{-1} = Q^{-1}P^{-1}.$$

If every term of a path $P$ is on a flat $S$ we say that $P$ is on $S$.

A path $P$ in which the terms are all distinct is called *simple*. As an example of a theorem on paths we offer the following.

4.33 *Let $P$ be a path from $X$ to $Y$ on a flat $S$ of $M$. Then there is a simple path from $X$ to $Y$ on $S$.*

PROOF: Let $P_1$ be a shortest path from $X$ to $Y$ on $S$. If it is not simple we can write it as a product $QRT$ where $R$ is re-entrant and nondegenerate. But then $QT$ is a path from $X$ to $Y$ on $S$ which is shorter than $P_1$. We conclude that in fact $P_1$ is simple.

The fundamental theorem about linear subclasses runs as follows.

4.34 *Let $C$ be a linear subclass of $M$. Let $S$ be a connected flat of $M$, and let $X$ and $Y$ be points on $S$ such that $Y$ is not in $C$. Then there is a simple path $P$ from $X$ to $Y$ on $S$ such that no term of $P$ other than $X$ belongs to $C$.*

PROOF: If possible choose $S$, $X$ and $Y$ so that the theorem fails and $dS$ has the least value consistent with this. Clearly $dS > 1$ and $Y$ is distinct from $X$. Figure 4B illustrates the following argument for the case $dS = 3$.

By 4.25 and 4.26 there is a connected $(dS-2)$-flat $U$ and two connected $(dS-1)$-flats $V$ and $W$ such that $X \subseteq U = \langle V \cap W \rangle$ and $V \cup W = S$. Moreover $Y$ is not on $V$ or $W$, by the choice of $S$, $X$, and $Y$.

By 4.27 there is a line $L$ on $S$ such that $\langle L \cap U \rangle = \phi$. It meets $V$ and $W$ in distinct points $T$ and $Z$, respectively. They are not both in $C$, for otherwise $Y$ would be in $C$. Without loss of generality we may suppose $T$ not in $C$,

By the choice of $S$, $X$, and $Y$ there is a simple path $Q$ from $X$ to $T$ on $V$ such that no term of $Q$ other than the first is a point on $C$. The product $Q(T, Y)$ may be taken as the required simple path $P$. As this result is contrary to the choice of $S$, $X$, and $T$ the theorem follows.

A slight modification of the above proof shows that we can arrange that $s(P) = dS$.

The following theorem is analogous to a well-known result in graph theory.

4.35 *Let $S$ be any flat of $M$. Then $S$ is connected if and only if for any two points $X$ and $Y$ on $S$ there exists a simple path on $S$ from $X$ to $Y$.*

PROOF: Suppose $S$ is connected. Let $X$ and $Y$ be points of $S$. Then there is a simple path on $S$ from $X$ to $Y$, by 4.34, with $C = \phi$.

Now suppose $S$ has a nontrivial separator $T$. There are points $X$ and $Y$ on $S$ such that $X \subseteq T$ and $Y \subseteq S - T$. If there is a path on $S$ from $X$ to $Y$ there are consecutive terms $X'$ and $Y'$ of $P$ such that $X' \subseteq T$ and $Y' \subseteq S - T$. But this is impossible since $X' \cup Y'$ is not a connected line, by 4.24.

4.36 *Let $C$ be any linear subclass of $M$. Let $S$ be a $d$-flat of $M$ on a $(d+1)$-flat $T$. Suppose all the points on $S$ and at least one other point on $T$ belong to $C$. Then all the points on $T$ belong to $C$.*

PROOF: Suppose the theorem fails. Then we can find points $X \epsilon C$ and $Y \notin C$, each being on $T$ but not on $S$. The flat $X \cup Y$ is connected since otherwise $S \subset S \cup X \subset T$, contrary to 4.14.

By 4.34 there is a path from $X$ to $Y$ on $X \cup Y$ whose second term, $X'$ say, is not in $C$. By 4.16 the connected line $X \cup X'$ has a point $X''$ in common with $S$. But $X''$ is not in C, by the definition of a linear subclass. This is contrary to hypothesis.

4.37 *Let $S$ be a connected flat of $M$. Let $C_1$ and $C_2$ be linear subclasses of $M$, neither of which includes all the points on $S$. Then there is a point $Z$ on $S$ which belongs to neither $C_1$ nor $C_2$.*

PROOF: Choose points $X$ and $Y$ on $S$ such that $X \notin C_1$ and $Y \notin C_2$. We may suppose $X$ and $Y$ distinct since otherwise the theorem is trivially true.

By 4.34 there is a point $X'$ on $S$ which is adjacent to $X$ and not in $C_2$. We may suppose $X \epsilon C_2$ and $X' \epsilon C_1$, for otherwise one of these points could be taken as $Z$. But now there is a third point $Z$ on the connected line $X \cup X'$, and this can belong neither to $C_1$ nor to $C_2$.

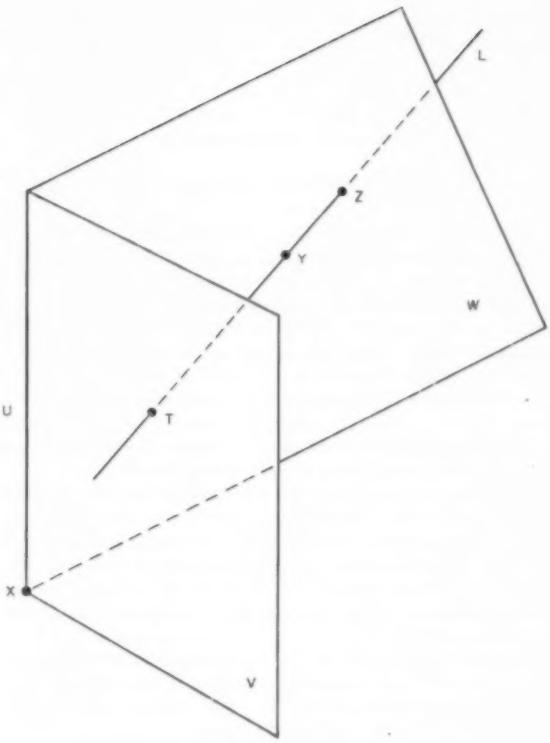



FIGURE 4B

As important special cases of 4.37 we note the following.

4.371 *Let $a$ and $b$ be distinct cells of a connected matroid $M$. Then there is an atom of $M$ which includes both $a$ and $b$.*

4.372 *Let $S$ and $T$ be subsets of $E$, each cut by some atom of $M$, $M$ being connected. Then some atom of $M$ cuts both $S$ and $T$.*

To prove 4.371 we take $C_1$ to be the set of points on $\langle E - \{a\} \rangle$, and $C_2$ to be the set on $\langle E - \{b\} \rangle$. 4.372 follows from 4.31.

When we apply 4.371 to the polygon-matroid of a graph we find that if $a$ and $b$ are distinct edges of a nonseparable graph $G$, then there is a polygon of $G$ through both $a$ and $b$. This is one of the theorems of Hassler Whitney on graphs.

## 4.4. Carriers

Let $M$ be a matroid on a set $E$, and $S$ be any subset of $E$. We proceed to discuss the relations between the flats of $M \cdot S$ and those of $M$. Dimensions referring to $M \cdot S$ will be distinguished in formulae by primes.

A *carrier* of $M \cdot S$ in $M$ is a subset $Z$ of $E$ such that

$$(M \times Z) \cdot S = M \cdot S$$

and $d\langle Z\rangle$ has the least value consistent with this property. Such a carrier exists, since $(M\times E)\cdot S = M\cdot S$.

4.41 *Let $Z$ be a carrier of $M\cdot S$ in $M$. Then* $\mathrm{dZ}=\mathrm{d'S}$.

PROOF: Assume there is an atom $X$ of $M$ such that $X\subseteq Z-S$. Choose $a\epsilon X$ and write

$$M_1=(M\times(Z-\{a\}))\cdot S.$$

Let $Y$ be an atom of $M_1$. Then there is an atom $U$ of $M$ such that $U\subseteq Z-\{a\}$ and $Y=S\cap U$. Hence there is an atom $Y'$ of $M\cdot S$ such that $Y'\subseteq Y$.

Conversely let $Y'$ be an atom of $M\cdot S=(M\times Z)\cdot S$. There is an atom $V$ of $M\times Z$ such that $Y'=S\cap V$. Suppose $a\epsilon V$. Then by applying Axiom II to $V$ and $X$ we find that there is an atom $U$ of $M\times Z$ such that $S\cap U$ is a non-null subset of $Y'$, and $a\notin U$. If $a\notin V$ we write $U=V$. In each case $U$ is an atom of $M\times(Z-\{a\})$ and $S\cap U$ is non-null. Hence there exists an atom $Y$ of $M_1$ such that $Y\subseteq S\cap U\subseteq Y'$.

Applying Axiom I to the foregoing results we find that $M_1=M\cdot S$. But this is contrary to the definition of $Z$ as a carrier of $M\cdot S$. We deduce that $d(Z-S)=-1$. The theorem now follows from 3.54.

4.42 *Let $Z$ be a carrier of $M\cdot S$ in $M$. Let $X$ be an atom of $M\cdot S$. Then there is just one atom $Y$ of $M\times Z$ such that $S\cap Y=X$.*

PROOF: Such an atom $Y$ exists, since $M\cdot S=(M\times Z)\cdot S$. Suppose $Y'$ is another such atom of $M\times Z$. Choose $a\epsilon S\cap Y=S\cap Y'=X$. By Axiom I we can find $b\epsilon Y'-Y$. By Axiom II there is an atom $U$ of $M\times Z$ such that $b\epsilon U\subseteq(Y\cup Y')-\{a\}$.

We have $S\cap U\subset S\cap Y=X$. Hence $S\cap U$ is null, since $M\cdot S$ satisfies Axiom I. Hence $U\subseteq Z-S$. But then $d\langle Z-S\rangle>-1$, and $dZ>d'S$, by 3.54. This is contrary to 4.41.

We can now define a mapping $\theta$ of the set of flats of $M\cdot S$ onto a class of flats of $M\times Z$, in the following way

(i) If $T$ is the null flat of $M\cdot S$ then $\theta T$ is the null flat of $M\times Z$.

(ii) If $T$ is an atom of $M\cdot S$, then $\theta T$ is the unique atom $U$ of $M$ on $Z$ such that $S\cap U=T$. (4.42).

(iii) If $T$ is a $d'$-flat of $M\cdot S$, where $d'\geqslant 1$, then $\theta T$ is the union of all atoms $\theta X$ of $M\times Z$ such that $X$ is an atom of $M\cdot S$ on $T$.

4.43 *If $T$ and $U$ are flats of $M\cdot S$, then $\theta T\subseteq\theta U$ if and only if $T\subseteq U$.*

This follows at once from the definition of $\theta$. It implies that $\theta T$ *and* $\theta U$ are distinct if $T$ and $U$ are distinct, and that $\theta T\subset\theta U$ if and only if $T\subset U$.

4.44 *If $T$ is a flat of $M\cdot S$, then $d'T=d\theta T$.*

PROOF: By 4.14 we can construct a sequence

$$T_{-1}, T_0, T_1, \ldots, T_{d'})$$

of flats of $M\cdot S$, where $d'=d'S$, such that each member of the sequence except $T_{d'}$ is a proper subset of its successor, such that $d'T_i=i$ for each $T_i$, and such that $T$ is a member of the sequence. Then $T=T_k$, where $k=d'T$.

By 4.14 and 4.43 the dimension is strictly increasing in the sequence

$$(\theta T_{-1}, \theta T_0, \theta T_1, \ldots, \theta T_{d'})$$

of flats of $M\times Z$. But $d\theta T_{-1}=-1$ and $d\theta T_{d'}=d'$, by 4.41. Hence $d\theta T_i=i$ for each member of the sequence. In particular $d\theta T=d'T$.

We have shown that $\theta$ is a $1-1$ mapping of the set of flats of $M\cdot S$ onto a class of flats of $M\times Z$, and that $\theta$ preserves dimension and inclusion relations. We may say therefore that $\theta$ transforms each geometrical figure made up of flats of $M\cdot S$ into a geometrically equivalent figure made up of flats of $M\times Z$.

Not every flat of $M\times Z$ need be of the form $\theta T$ however. A simple example is provided by a matroid of three cells $a$, $b$, and $c$, and three atoms $\{a, b\}$, $\{b, c\}$, and $\{c, a\}$. We put $S=\{a, b\}$ and $Z=\{a, b, c\}$. *Then* $Z$ is clearly a carrier of $M\cdot S$. But the set $S$ is a flat of dimension 2 in $M\cdot S$ and dimension 1 in $M\times Z$. Hence $S$ cannot be of the form $\theta T$, with $T$ a flat of $M\cdot S$.

In general we have the following theorem

4.45 *If $V$ is a flat of $M\times Z$, then $S\cap V$ is a flat of $M\cdot S$, and moreover $V\subseteq\theta(S\cap V)$.*

PROOF: The first assertion follows from 1.11 and the definition of $M\cdot S$.

It is evident that $S\cap V\subseteq\theta(S\cap V)$.

Suppose $b\epsilon V-S$. There is an atom $X_b$ of $M\times Z$ such that $b\epsilon X_b$ and $S\cap X_b\subseteq S\cap V$. Choose such an $X_b$ so that $|S\cap X_b|$ has the least possible value. Then $S\cap X_b$ is non-null, by 4.41 and 3.54.

Assume $S\cap X_b$ is not an atom of $M\cdot S$. There exists $Y\epsilon M\cdot S$ such that $Y\subset S\cap X_b$. Then $b$ is not a cell of $\theta Y$, by the choice of $X_b$. Choose $a\epsilon Y$. Applying Axiom II to $X_b$ and $\theta Y$ we find that $M\times Z$ has an atom $Y'$ such that $b\epsilon Y'\subseteq(X_b\cup Y)-\{a\}$. But then $S\cap Y'\subset S\cap X_b$, and the definition of $X_b$ is contradicted.

We deduce that $S\cap X_b$ is an atom of $M\cdot S$. Hence $\theta(S\cap X_b)=X_b$, by 4.42, and therefore $b\epsilon\theta(S\cap V)$. Since this is true for all $b\epsilon V-S$ we have $V-S\subseteq\theta(S\cap V)$.

Combining the above results we find that $V\subseteq\theta(S\cap V)$.

## 5. Specializations

### 5.1. Matroids and Chain-Groups

As before we suppose given a commutative ring $R$ with a unit element and no divisors of zero.

We have to consider the following problem. Given a matroid $M$ we are to determine whether there exists a chain-group $N$ over $R$ such that $M=M(N)$. In this work we solve the problem only in some special cases, each time with the help of the following theorem.

5.11 *Let $M$ be a matroid on a set $E$. With each atom $X$ of $M$ let there be associated a chain $f(X)$ on $E$ over $R$ such that $||f(X)||=X$. Suppose further that whenever $X$, $Y$, and $Z$ are distinct points on the same line of $M$ there exist non-zero elements $r$, $s$, and $t$ of $R$ such that*

$$rf(X)+sf(Y)+tf(Z)=0$$

*Let $N$ be the chain-group on $E$ over $R$ generated by the chains $f(X)$.*

*Then $M = M(N)$.*

PROOF: Let $T$ be any point of $M(N)$. There is an elementary chain $f$ of $N$ such that $\|f\| = T$. By the definition of $N$ we can find a flat $S$ of $M$ with the following properties.

(i) There is a nonzero element $r$ of $R$ such that $rf$ is a sum of chains $tf(X)$ with $X \subseteq S$, $t \in R$.

(ii) $dS$ has the least value consistent with (i).

If possible choose $a \in S - T$ and write $S' = \langle S - \{a\} \rangle$. The set $C$ of all points of $M$ which are subsets of $S'$ is a linear subclass of $M$. (4.3).

If $g$ and $h$ are chains of $N$ we write

$$g \equiv h \mod S'$$

to denote that $g - h$ is a sum of chains $tf(X)$ such that $X \subseteq S'$. In particular we may write

$$rf = \sum_i t_i f(Y_i) \mod S', \qquad t_i \in R. \tag{1}$$

where the $Y_i$ are points of $M$ such that $a \in Y_i \subseteq S$,

Let $Y$ be any point of $M$ such that $a \in Y \subseteq S$. For each $Y_i$ the flat $Y \cup Y_i$ of $M$ is connected, by 3.48. Hence there is a path $P_i$ from $Y_i$ to $Y$ on $S$ such that no term of $P_i$ is a point of $C$, by 4.34.

If $U$ and $V$ are consecutive points in $P_i$ then the line $U \cup V$ meets $S'$ in a point $W$ of $C$, by 4.16. Then by hypothesis we have

$$x_i f(U) \equiv y_i f(V) \mod S',$$

where $x_i$ and $y_i$ are nonzero members of $R$. If $Y_i \neq Y$ we state this congruence for each pair of consecutive terms of $P_i$ and combine the results. In every case we have

$$r_i f(Y_i) \equiv s_i f(Y) \mod S', \tag{2}$$

where $r_i$ and $s_i$ are nonzero elements of $R$. Multiplying (1) by the product of the elements $r_i$ and using (2) we find that

$$r' f \equiv s' f(Y) \qquad \mod S',$$

where $r'$ and $s'$ are elements of $R$, $r'$ being nonzero.

Since $a$ does not belong to the domain of $f$ we deduce that $s' = 0$, which is contrary to the definition of $S$. We conclude that in fact $S = T$. Hence there is a point $T'$ of $M$ on $S$ such that $T' \subseteq T$.

Conversely let $T'$ be any point of $M$. Then $f(T')$ is a chain of $N$. Hence there is a point $T$ of $M(N)$ such that $T \subseteq T'$.

Applying Axiom I to these two results we find that the matroids $M$ and $M(N)$ are identical.

## 5.2. Representative Matrices

It is usually convenient to specify a chain-group over $R$ in the following way.

First we enumerate the set $E$, supposed non-null, as $\{e_1, e_2, \ldots, e_n\}$. For each chain $f$ on $E$ over $R$ we define the *representative vector* of $f$ as the 1-rowed matrix.

$$\{f(e_1), f(e_2), \ldots, f(e_n)\}.$$

Clearly linear relations between chains hold also for their representative vectors, and conversely.

A matrix $K$, with elements in $R$, is called a *representative matrix* of $N$ if it satisfies the following conditions.

(i) The rows of $K$ are linearly independent representative vectors of chains of $N$.

(ii) Every nonzero chain of $N$ has a representative vector which is a linear combination of rows of $K$.

Linear dependence is of course defined in terms of the elements of $R$ as coefficients.

A matrix $K$ satisfying (i) and (ii) completely determines the chain group $N$. It is also evident that any matrix $K$ with linearly independent rows can be interpreted as a representative matrix of a chain-group.

Condition (i) ensures that no row of $K$ consists solely of zero elements. If $N$ has only a zero chain it is convenient to say it has a *null* representative matrix, with no rows but $n$ columns. Such a matrix is considered to satisfy (i).

Let $K$ be a representative matrix of $N$ and let $S$ be any subset of $E$. We define $K(S)$ as the submatrix of $K$ made up of those columns which correspond to cells of $S$. If $S$ is null then $K(S)$ is a second kind of *null* matrix, being without columns but having as many rows as $K$. If the number of such rows is nonzero the matrix $K(S)$ is not regarded as having linearly independent rows.

As an immediate consequence of the definitions we have

5.21 *Let $K$ be a representative matrix of $N$ and let $S$ be a subset of $E$ such that the rows of $K(S)$ are linearly independent. Then $K(S)$ is a representative matrix of $N \cdot S$.*

Suppose the representative matrix $K$ of $N$ has $r$ rows and $n$ columns. It may happen that there is a subset $S$ of $E$, satisfying $|S| = r$ such that $K(S)$ is a unit matrix, that is having only 1's in the main diagonal and only 0's elsewhere. In this case $S$ is evidently a dendroid of $N$. We call $K$ a *standard* representative matrix of $N$ associated with the dendroid $S$. If we wish we may think of the unit matrix $K(S)$ as occupying the first $r$ columns of $K$. For this can be arranged by adjusting the initial enumeration of $E$.

5.22 *Let $K$ be a standard representative matrix of $N$ associated with a dendroid $D$. Then the rows of $K$ are representative vectors of elementary chains of $N$, and these chains constitute a dendroid-basis of $N$ corresponding to $D$.*

PROOF: By 2.3 there is a dendroid basis $\{f_a | a \in D\}$ of $N$. The theorem follows from the Corollary to 2.31.

A null representative matrix of a trivial chain-group $N$ may be regarded as standard, being associated with a null dendroid.

Let $K$ be a standard representative matrix of $N$,

associated with a dendroid $D$ and having $r$ rows and $n$ columns.

We construct a matrix $K^*$, having $n-r$ rows and $n$ columns, in the following way. The submatrix of $K^*$ occupying the $n-r$ columns corresponding in position to those of $K(E-D)$ in $K$ is a unit matrix. The other columns constitute a submatrix of $K^*$ equal to minus the transpose of $K(E-D)$. We illustrate the construction by an example, in which $R$ is the ring of integers

$$K=\left\{\begin{array}{ccccc} \overbrace{1 & 0}^{D} & \overbrace{2 & -1 & 9}^{E-D} \\ 0 & 1 & 0 & 5 & 6 \end{array}\right\}$$

$$K^*=\left\{\begin{array}{ccccc} -2 & 0 & 1 & 0 & 0 \\ 1 & -5 & 0 & 1 & 0 \\ -9 & -6 & 0 & 0 & 1 \end{array}\right\}$$

5.23 *$K^*$ is a standard representative matrix of $N^*$ associated with the dendroid $E-D$.*

PROOF: $E-D$ is a dendroid of $N^*$ by 2.43. The theorem follows from 2.42 and 5.22.

Returning to the matrix $K$ we consider an arbitrary submatrix $A$ of $K(E-D)$. Let $A_1$ be the submatrix of $K$ constituted by those rows which meet $A$. Let $A_2$ be the submatrix of $A_1$ constituted by those of its columns which have 1's in $K(D)$. Thus $A_2$ is a unit matrix having the same number of rows as $A$. We refer to $A_2$ as the *annex* Ann $(A)$ of $A$. The submatrix of $K$ made up of $A$ and Ann $(A)$ is the *extension* Ext $(A)$ of $A$. We give an example below

$$\begin{array}{c} \overbrace{\phantom{1\ 0\ 0\ 0\ 0}}^{D}\ \overbrace{\phantom{1\ -1\ 1\ 1\ -1}}^{E-D} \\ \left(\begin{array}{ccccc|ccccc} 1 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 1 & -1 \\ 0 & 1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 \end{array}\right) \end{array}$$

Ann $A$ ↗ … ↖ $A$

FIGURE 5A.

It is not of course essential that the rows and columns of $A$ should occur consecutively in $K$.

5.24 *Let $A$ be any submatrix of $K(E-D)$. Then* Ext $(A)$ *is a standard representative matrix of a minor of $N$. The corresponding dendroid $D_1$ is made up of those cells which correspond to columns of* Ann $(A)$.

PROOF: Let $D_1$ and $Z$ be subsets of $E$ made up of the cells which correspond to columns of Ann $(A)$ and $A$ respectively. Let $B$ be the submatrix of $K(E-D)$ made up of those rows of $K(E-D)$ which meet $A$. Thus Ann $(B)$ = Ann $(A)$.

We note that the rows of Ext $(B)$ and Ext $(A)$ are linearly independent. For each of these matrices contains the unit matrix Ann $(A)$ and has the same number of rows as Ann $(A)$.

It is clear that a chain $f$ of $N$ has a restriction in $N\times((E-D)\cup D_1)$ if and only if its representative vector is a linear combination of rows of $A_1$. Hence Ext $(B)$ is a representative matrix of $N\times((E-D)\cup D_1)$, and therefore Ext $(A)$ is a representative matrix of the minor.

$$N_1=(N\times((E-D)\cup D_1))\cdot(Z\cup D_1),$$

by 5.21. Evidently Ext $(A)$ satisfies the definition of a standard representative matrix of $N_1$ associated with $D_1$.

A minor of $N$ having a representative matrix of the form Ext $(A)$, where $A$ is a submatrix of $K(E-D)$, will be said to be *visible* in $K$.

In general we cannot assert that there is a standard representative matrix of $N$ associated with each dendroid. However we have the following theorem.

5.25 *Suppose that either $R$ is a field or that $N$ is regular. Then if $D$ is any dendroid of $N$ we can find a standard representative matrix of $N$ associated with $D$.*

PROOF: There is a dendroid-basis $\{f_a \mid a\epsilon D\}$ of $N$. If $N$ is regular we can take each chain $f_a$ to be primitive. So in all cases covered by the theorem we may suppose $f_a(a)$ to have a reciprocal in $R$. Multiplying each chain $f_a$ by the reciprocal of $f_a(a)$ we obtain a dendroid-basis $\{g_a \mid a\epsilon D\}$ of $N$ such that $g_a(a)=1$ for each $a\epsilon D$. Clearly there is a standard representative matrix of $N$, associated with $D$, whose rows are the representative vectors of the chains $g_a$.

We may now supplement 5.24 as follows.

5.26 *Suppose that either $R$ is a field or $N$ is regular. Let $L$ be any minor of $N$. Then we can find a standard representative matrix $K$ of $N$, associated with a dendroid $D$, such that $L$ is visible in $K$.*

PROOF: Let $D_1$ be any dendroid of $L$ and let $Z$ be the set of all cells of $L$ not in $D_1$. Then we may write

$$L=(N\times U)\cdot(Z\cup D_1),$$

by 3.243.

We recall that the dendroids of a chain-group $N$ are the dendroids of the corresponding matroid $M(N)$, by 2.3.

Let $V$ be a carrier of $M(L)$, that is $(M(N)\times U)\cdot(Z\cup D_1)$ by 3.31, in $M(N)\times U$. Then $D_1$, being a dendroid of $M(L)$, is also a dendroid of $M(N)\times V$. Hence $D_1$ is a dendroid of $N\times V$.

We have

$$L=(N\times V)\cdot(Z\cup D_1),$$

for the restrictions to $Z\cup D_1$ of the members of a dendroid basis of $N\times V$ corresponding to $D_1$ evidently form a dendroid-basis of $L$. (2.31, Corollary).

Let $D_2$ be a dendroid of $N\cdot(E-V)$, and write $D_1\cup D_2=D$. Then $D$ is a dendroid of $N$, by 3.31 and 3.53. Let $K$ be a standard representative matrix of $N$ associated with $D$. Such a matrix exists, by 5.25.

Let $f$ be any chain of $N\times(E-D_2)$. Its restriction to $E-V$ has a domain not meeting $D_2$, and is therefore zero. Since $E-D_2=(E-D)\cup D_1$ we deduce that

$$L=(N\times((E-D)\cup D_1))\cdot(Z\cup D_1)$$

But by the argument of 5.24 this minor is visible in $K$.

### 5.3. Characterization of Binary Matroids

A binary matroid is by definition the matroid of a binary chain-group $N$, that is a chain-group over the field $GF(2)$ of residues mod 2.

5.31 *Let $M$ be any binary matroid. Then $M^*$ and the minors of $M$ are binary.* (2.66 and 3.31).

A binary matroid is conveniently specified by giving a representative matrix of the associated binary chain-group.

5.32 *Let $N$ be a binary chain-group on a set $E$. Then every nonzero chain of $N$ is a sum of elementary chains of $N$ with disjoint domains.*

PROOF: If possible let $f$ be a nonzero chain of $N$ for which the theorem fails, and for which $\|f\|$ has the least number of cells consistent with this condition.

There is an elementary chain $g$ of $N$ such that $\|g\|\subseteq\|f\|$. But $g\neq f$, for otherwise $f$ would satisfy the theorem. Hence the chains $g$ and $f+g$ have disjoint domains, each with fewer cells than $\|f\|$. Accordingly $g$ and $f+g$ satisfy the theorem, and therefore $f$ satisfies it, contrary to assumption. The theorem follows.

In a binary chain-group there is little distinction between a chain and its domain. If $S_1, S_2, \ldots, S_k$ are subsets of $E$, not necessarily all distinct, we define their mod 2 sum as the subset $S$ of $E$ such that $a\epsilon S$ if and only if the number of suffices $i$, $1\leq i\leq k$, satisfying $a\epsilon S_i$ is odd. This addition of subsets is evidently commutative and associative, and we use the ordinary additive notation for it. Chains $f_i$ on $E$ over $GF(2)$ evidently satisfy

$$5.33\qquad \|\sum_{i=1}^{k} f_i\| = \sum_{i=1}^{k} \|f_i\|\cdot$$

We conclude this section with two theorems characterizing binary matroids.

5.34 *A matroid $M$ on $E$ is binary if and only if any non-null mod 2 sum of atoms of $M$ is a union of disjoint atoms of $M$.*

PROOF: Suppose $M=M(N)$, where $N$ is a binary chain-group. A non-null mod 2 sum of atoms of $M$ is the domain of a non-null chain of $N$, by 5.33. This is a union of disjoint atoms of $M$ by 5.32.

Conversely suppose $M$ satisfies the stated condition. Let $N$ be the chain-group generated by those chains on $E$ over $GF$ (2) whose domains are atoms of $M$. Clearly each atom of $M$ contains an atom of $M(N)$. On the other hand each atom of $M(N)$ is a mod 2 sum of atoms of $M$ and therefore contains an atom of $M$. Hence $M=M(N)$, by Axiom I, and so $M$ is binary.

5.35 *A matroid $M$ is binary if and only if each connected line of $M$ is on exactly three points.*

PROOF: Suppose $M$ is binary. Let $L$ be any connected line of $M$. There are at least three distinct points on $L$, by 4.23. Denote two of them by $X$ and $Y$, and let $Z$ be any other.

Since $X\cup Z=L=X\cup Y$, by 4.22, we have $Y-X\subseteq Z$. Similarly $X-Y\subseteq Z$. Hence $Z$ contains the mod 2 sum $X+Y$, which is non-null by Axiom I. But $X+Y$ is a union of disjoint atoms of $M$, by 5.34. Hence $Z=X+Y$, by Axiom I. Thus $L$ is on just one point other than $X$ and $Y$.

Conversely suppose there are just three points on each connected line of $M$. If $X$ is any point (atom) of $M$ we write $f(X)$ for the chain on $E$ over $GF(2)$ whose domain is $X$. Let $N$ be the binary chain-group on $E$ generated by the chains $f(X)$.

Let $X$, $Y$, and $Z$ be the three points on any connected line $L$ of $M$. Then $X+Y+Z$ is null, by 4.21. Hence $f(X)+f(Y)+f(Z)=0$, by 5.33. Accordingly $M=M(N)$, by 5.11, and so $M$ is binary.

### 5.4. Regular Chain-Groups and Matroids

A regular matroid is by definition the matroid of a regular chain-group $N$, as defined in section 1.2.

5.41 *Let $M$ be any regular matroid. Then $M^*$ and the minors of $M$ are regular.* (2.47, 2.66, 3.26 and 3.31).

Let $f$ and $g$ be two integral chains on a set $E$. We say that $f$ *conforms* to $g$ if $||f||\subseteq||g||$ and $f(a)g(a)>0$ for each $a\epsilon||f||$.

5.42 *Let $N$ be a regular chain-group on $E$, and let $f$ be a nonzero chain of $N$. Then there exists a primitive chain of $N$ conforming to $f$.*

PROOF: If possible choose $f$ so that the theorem fails and $\|f\|$ has the least number of elements consistent with this condition.

Since $N$ is regular it has a primitive chain $g$ such that $\|g\|\subseteq\|f\|$. Choose $a\epsilon\|g\|$ such that $f(a)$ has the least possible absolute value. Replacing $g$ by $-g$ if necessary we arrange that $g(a)\cdot f(a)>0$. Write

$$h=f-f(a)g(a)g$$

If $h=0$ then $f$ is a positive multiple of $g$, and so $g$ conforms to $f$. If $h\neq 0$ it is clear that $h$ conforms to $f$. Moreover $\|h\|$ then has fewer cells that $\|f\|$ and so there is a primitive chain $k$ of $N$ conforming to $h$. But then $k$ conforms also to $f$. So in each case the choice of $f$ is contradicted. The theorem follows:

5.43 *Let $N$ be a regular chain-group on $E$, and let $f$ be a nonzero chain of $N$. Then $f$ can be represented as a sum of primitive chains of $N$, each conforming to $f$.*

PROOF: Let $Z(k)$, where $k$ is a chain of $N$, denote the sum of the absolute values of the coefficients in $k$.

If possible choose $f$ so that the theorem fails and $Z(f)$ has the least value consistent with this. By 5.42 there is a primitive chain $g$ of $N$ conforming to $f$. Write $h=f-g$. If $h=0$ then $f=g$. If $h\neq 0$ it is clear that $Z(h)<Z(f)$ and that $h$ conforms to $f$. Hence $h$ is a sum of primitive chains of $N$, each conforming to $h$ and therefore to $f$. In either case the definition of $f$ is contradicted.

Let $f$ and $g$ be integral chains on $E$ and let $q$ be an integer not less than 2. We call $g$ a *$q$-representative* of $f$ if the following conditions are satisfied for each $a\epsilon E$.

(i) $g(a)\equiv f(a) \bmod q$,
(ii) $|g(a)|<q$.

5.44 *Let $f$ be a chain of a regular chain-group $N$ on $E$. Then for each integer $q \geq 2$ some chain of $N$ is a $q$-representative of $f$.*

PROOF: There is at least one chain $g$ of $N$ which satisfies condition (i), namely $f$. For each such $g$ we define $Y(g)$ as the number of cells $a \epsilon E$ satisfying $|g(a)| \geq q$. We choose a particular $g$ satisfying (i) so that $Y(g)$ has the least possible value.

If $Y(g) > 0$ choose $b \epsilon E$ so that $|g(b)| \geq q$. By 5.43 there is a primitive chain $h$ of $N$ conforming to $g$ and satisfying $h(b) = \pm 1$. Write

$$g_1 = g - qh.$$

Clearly $g_1$ satisfies (i). Moreover we have

$$|g_1(b)| < |g(b)|,$$

$$(|g(a)| < q) \rightarrow (|g_1(a)| < q).$$

Hence $Y(g_1) \leq Y(g)$, with equality only if $|g_1(b)| \geq q$.

If $|g_1(b)| \geq q$ we repeat the process with $g_1$ replacing $g$, and with the same choice of $b$. Proceeding in this way we eventually obtain a chain $g'$ of $N$ such that $g'$ satisfies (i) and $Y(g') < Y(g)$. But this is contrary to the choice of $g$. We deduce that in fact $Y(g) = 0$. The theorem is thus true.

We use 5.44 to prove the following important theorem about regular matroids.

5.45 *Every regular matroid is binary.*

PROOF: Let $M$ be a regular matroid. Then $M = M(N)$, where $N$ is a regular chain-group. For each chain $f$ of $N$ let $f'$ be the chain over $GF(2)$ derived from it by replacing each coefficient in $f$ by its residue mod 2. Then the chains $f'$ constitute a binary chain-group $Q$.

Let $X$ be any atom of $M$. It is the domain of a primitive chain of $N$, and therefore of a nonzero chain of $Q$. There is an atom $Y$ of $M(Q)$ such that $Y \subseteq X$.

Conversely let $Y$ be an atom of $M(Q)$. There is a chain $f$ of $N$ such that $Y$ is the domain of $f'$. But there is a chain $g$ of $N$ which is a 2-representative of $f$, by 5.44, and we have $\|g\| = \|f'\| = Y$. Hence there is an atom $X$ of $M = M(N)$ such that $X \subseteq Y$.

Applying Axiom I we find that $M = M(Q)$. Thus $M$ is binary.

There are some important properties of regular chain-groups which are expressible in terms of their standard representative matrices.

5.46 *Let $N$ be a regular chain-group on $E$. Let $D$ be any dendroid of $N$. Let $K$ be a standard representative matrix of $N$ associated with $D$, and let $S$ be any subset of $E$ such that $|S| = |D|$. Then the determinant of the square submatrix $K(S)$ is $\pm 1$ if $S$ is a dendroid of $N$, and zero otherwise.*

PROOF: It is clear that $S$ meets the domain of every nonzero chain of $N$ if and only if $K(S)$ is nonsingular. So, by 2.23 det $K(S)$ is nonzero if and only if $S$ is a dendroid of $N$.

Suppose $K(S)$ is nonsingular. Let $K_1$ be the standard representative matrix of $N$ associated with $S$ (5.25). Since the rows of $K_1$ are linear combinations of those of $K$ there is a square matrix $A$ of integers such that $AK = K_1$. But then $AK(S) = K_1(S)$. On taking determinants we have

$$\det A \cdot \det K(S) = 1,$$

since $K_1(S)$ is a unit matrix. Since $A$ and $K(S)$ are matrices of integers it follows that det $K(S) = \pm 1$.

Using the formula for the determinant of a product of two rectangular matrices we can deduce from 5.46 that the number of distinct dendroids of $N$ is

$$\det (KK^T),$$

where $K^T$ is the transpose of $K$. This generalizes a well-known formula for the number of spanning trees in a graph $G$.

There is a converse of 5.46 to the effect that if $K$ is an $r$-rowed matrix of integers whose rows are linearly independent, and in which the determinants of the $r \times r$ submatrices are restricted to the values 1, $-1$, and 0, then $K$ is a representative matrix of a regular chain group. We do not use this theorem in what follows, and we leave its proof as an exercise to the reader.

The following consequence of 5.46 is of interest.

5.47 *Let $N$ be a regular chain-group on $E$. Let $K$ be a standard representative matrix of $N$ associated with a dendroid $D$. Then the determinants of the square submatrices of $K(E-D)$ are restricted to the values 1, $-1$ and 0.*

PROOF: Let $A$ be such a square submatrix. Let $T$ be the subset of $E$ corresponding to the columns of $K(E-D)$ meeting $A$ and those of $K(D)$ not meeting Ann $(A)$.

Now $K(T)$ is square, and det $K(T) = 1, -1$ or 0 by 5.46. But expansion of det $K(T)$, using the columns in $K(D)$, shows that det $K(T) = \pm$ det $A$. The theorem follows.

We can express this result by saying that $K(E-D)$ is *completely unimodular*. Evidently $K$ is itself completely unimodular.

## 5.5. Some Binary Matroids Which Are Not Regular

In this section we show that the converse of 5.45 is not true. We do so by using the following observation.

5.51 *Let $K$ be a standard representative matrix, associated with a dendroid $D$, of a binary chain-group $N$ on $E$. Then if $M(N)$ is regular we can replace each zero element of $K$ by the integer 0, and each nonzero element by an integer $+1$ or $-1$ in such a way as to transform $K(E-D)$ into a completely unimodular matrix.*

PROOF: If $M$ is regular we can write $M = M(N) = M(N_1)$, where $N_1$ is a regular chain-group.

Each row of $K$ is the representative vector of an elementary chain of $N$. We replace it by the representative vector of a primitive chain of $N_1$ with the same domain. Performing this operation for each row, and then making appropriate multiplications of rows by $-1$ we evidently transform $K$ into a standard representative matroid $K_1$ of $N_1$ associated with $D$.

We observe that $K$ is transformed into $K_1$ by an operation of the kind specified, and that $K(E-D)$ is transformed into the completely unimodular matrix $K_1(E-D)$, by 5.47.

We now investigate a binary chain-group $N$ with the standard representative matrix shown below

$$\begin{Bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{Bmatrix}$$

We shall describe a binary matroid corresponding to such a chain-group as being "of type BI".

5.52 *No matroid of Type BI is regular.*

PROOF: Suppose the theorem false. Then by 5.51 there is a completely unimodular matrix $A$ of the following form.

$$\begin{Bmatrix} x_{11} & 0 & x_{13} & x_{14} \\ x_{21} & x_{22} & 0 & x_{24} \\ 0 & x_{32} & x_{33} & x_{34} \end{Bmatrix}$$

Here each $x_{ij}$ is $+1$ or $-1$.

The determinant of the square submatrix

$$\begin{Bmatrix} x_{11} & x_{14} \\ x_{21} & x_{24} \end{Bmatrix}$$

is clearly even. It is therefore zero since $A$ is completely unimodular. We deduce that $x_{11}x_{24}=x_{14}x_{21}$, that is

$$x_{11}x_{21}x_{14}x_{24}=1. \tag{1}$$

Similarly we have

$$x_{22}x_{32}x_{24}x_{34}=1, \tag{2}$$

$$x_{13}x_{33}x_{14}x_{34}=1. \tag{3}$$

We have also

$$\begin{vmatrix} x_{11} & 0 & x_{13} \\ x_{21} & x_{22} & 0 \\ 0 & x_{32} & x_{33} \end{vmatrix} = x_{11}x_{22}x_{33}+x_{13}x_{21}x_{32}.$$

But this number is even, and therefore zero by the complete unimodularity of $A$. Hence

$$x_{11}x_{21}x_{22}x_{32}x_{13}x_{33}=-1 \tag{4}$$

But equations (1), (2), (3), and (4) are inconsistent, for they imply

$$\begin{aligned} -1 &= (x_{11}x_{21}x_{22}x_{32}x_{13}x_{33})x_{14}^2x_{24}^2x_{34}^2 \\ &= (x_{11}x_{21}x_{14}x_{24})(x_{22}x_{32}x_{24}x_{34})(x_{13}x_{33}x_{14}x_{34}) \\ &= 1. \end{aligned}$$

The theorem follows.

The dual of a matroid of Type BI will be said to be "of Type BII". Such a matroid is binary but not regular, by 5.31, 5.41, and 2.65.

By another application of 5.31 and 5.41 we have

5.53 *Let $M$ be a binary matroid having a minor of Type BI or BII. Then $M$ is not regular.*

Our next main task is the proof of the converse of this theorem, that if a binary matroid is not regular it must have a minor of Type BI or BII. Our proof will be based on 5.11. But the question whether it is possible to assign suitable integral chains $f(X)$ to the atoms X of a given binary matroid raises grave difficulties. Apparently these can only be resolved by the use of the homotopy theory developed in the next Chapter.

## 5.6. Graphic and Cographic Matroids

We have discussed specializations from general matroids to binary matroids, and from binary matroids to regular ones. The next step in this progression brings us to the graphic and cographic matroids.

We have defined a *graphic* matroid as one which can be interpreted as the bond-matroid of a graph (sec. 2.5). A *cographic* matroid is one which can be interpreted as the polygon-matroid of a graph.

5.61 *The cographic matroids are the duals of the graphic matroids.* (2.67).

5.62 *Let $M$ be any graphic (cographic) matroid. Then the minors of $M$ are graphic (cographic).* (3.321, 3.322, 3.371, 3.372).

5.63 *Every graphic or cographic matroid is regular.* (1.24, 5.41 and 5.61).

The last of these theorems shows that the graphic and cographic matroids can be regarded as specializations of the regular ones. We shall therefore have to discuss the question: when is a regular matroid graphic? Here we shall only state the result, leaving the proof for a later chapter.

A complete *5-graph*, or *5-clique* is a graph having just five vertices $a_1, \ldots, a_5$ and just ten edges $L_{ij}$, the ends of $L_{ij}$ being $a_i$ and $a_j$ $(1 \leq i < j \leq 5)$. A *Thomsen graph*, or *4-cage* is a graph having just six vertices $a_1, a_2, a_3, b_1, b_2, b_3$ and just nine edges $L_{ij}$, the ends of $L_{ij}$ being $a_i$ and $b_j$. $(1 \leq (i, j) \leq 3)$. These two graphs are illustrated in figure 5B.

These two graphs are often spoken of as the *Kuratowski graphs*, in view of Kuratowski's Theorem that every nonplanar graph contains a subdivision of one of them.

We shall prove in a later chapter that a regular matroid $M$ is nongraphic if and only if one of its minors

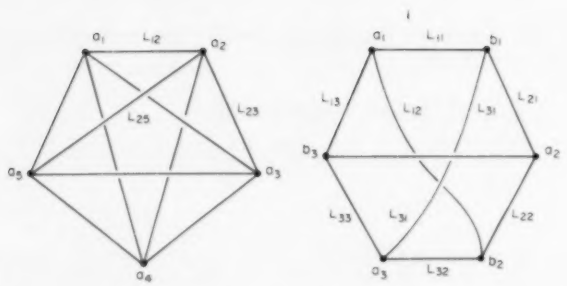



FIGURE 5B

is the polygon-matroid of a Kuratowski graph. Dually $M$ is noncographic if and only if one of its minors is the bond-matroid of a Kuratowski graph.

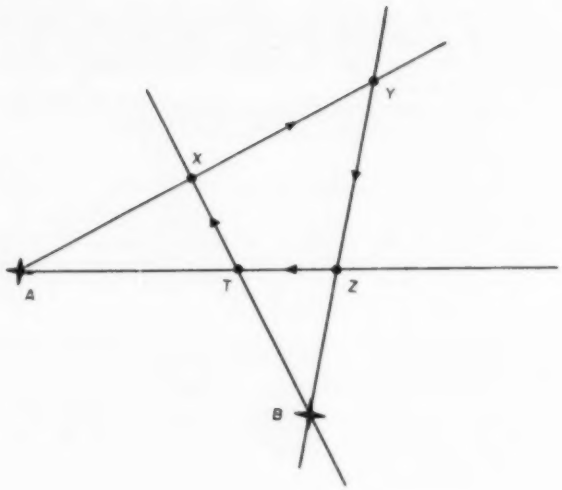



FIGURE 6A

## 6. Homotopy

### 6.1. Elementary Re-entrant Paths

Throughout this chapter we suppose given a matroid $M$ and a linear subclass $C$ of $M$. We describe a path in $M$ as being *off* $C$ if it has no term which is a vertex of $C$, and we study the properties of those re-entrant paths in $M$ which are off $C$. As a special case of course we can put $C=\phi$ and study the set of all re-entrant paths of $M$.

We specify four kinds of re-entrant path off $C$ as fundamental, or *elementary*, and we show that every re-entrant path off $C$ can, in a sense to be explained, be expressed as a combination of elementary ones.

The *first* kind of elementary re-entrant path consists of all paths off $C$ of the form $(X, Y, X)$. The *second* consists of all paths off $C$ of the form $(X, Y, Z, X$ where $d(X\cup Y\cup Z)\leqslant 2$.

Suppose $P$ is a plane of $M$ on which there are two distinct points $A$ and $B$ of $C$ such that each connnected line on $P$ is on either $A$ or $B$. Then any path, on $P$ and off $C$, of the form $(X, Y, Z, T, X)$ where $X$, $Y$, $Z$, and $T$ are distinct, $X\cup Y$ and $Z\cup T$ are lines on $A$, and $Y\cup Z$ and $T\cup X$ are lines on $B$, is an elementary re-entrant path of the *third* kind.

Such a path is shown in figure 6A. In this and other diagrams we adopt the convention that points of $C$ are to be represented by 4-pointed stars.

Now let $J$ be a 3-flat of $M$ on which there are three points $Z_1$, $Z_2$, and $Z_3$ such that $Z_1\cup Z_2$, $Z_2\cup Z_3$ and $Z_3\cup Z_1$ are disconnected lines. Let there be just six connected planes on $J$, two on each of these lines. We enumerate the planes as $P_1, P_2, \ldots, P_6$ in such a way that

$$\langle P_i\cap P_{i+3}\rangle = Z_j\cup Z_k,$$

where $(i, j, k)$ is any permutation of $(1, 2, 3)$.

If $1\leqslant i<j\leqslant 6$ the flat $\langle P_i\cap P_j\rangle$ is a line, by 4.172. We denote it by $L_{ij}$. If $j=i+3$ then $L_{ij}$ is the disconnected line

$$(Z_1\cup Z_2\cup Z_3)-Z_i.$$

If $j\neq i+3$ let $k$ be that integer 1, 2, or 3 which is not congruent to $i$ or $j$ mod 3. Then $L_{ij}$ is on $Z_k$, and it meets $P_k$ and $P_{k+3}$ in two distinct points. Hence $L_{ij}$ is connected by 4.23. Clearly it is on no connected plane on $J$ other than $P_i$ and $P_j$.

We observe that $J=P_1\cup P_2$, by 4.14, and is thus connected, by 3.48. It follows that each connected line on $J$ is on two connected planes of $J$. Hence the 12 lines $L_{ij}$, $j\neq i+3$, are the only connected lines on $J$.

We write

$$X_{ijk}=\langle P_i\cap P_j\cap P_k\rangle$$

for $1\leqslant i<j<k\leqslant 6$. Then $X_{ijk}$ is a point on $J$, by 4.173, being identical with $\langle L_{ij}\cap P_k\rangle$. If two of the suffices $i, j$, and $k$ are congruent mod 3, then $X_{ijk}$ is one of the points $Z_1$, $Z_2$, and $Z_3$. The remaining eight points $X_{ijk}$ are all distinct, for on any one of them there can be only three planes such that each is on $L_{14}$, $L_{25}$, or $L_{36}$. These eight points, together with $Z_1$, $Z_2$, and $Z_3$ are the only points on $J$. For each point on $J$ is on three distinct connected planes of $J$, by two applications of 4.26.

The geometrical figure just described is shown in figure 6B. Three lines are broken to indicate disconnection.

We refer to this structure as a *box*, saying that the planes $P_i$ are its *sides*, the 12 connected lines are its *edges*, the points $Z_i$ are its *radiants* and the other eight points are its *corners*.

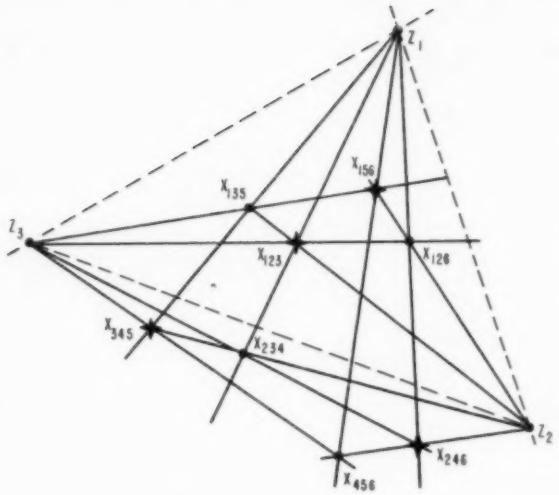



FIGURE 6B

It may happen that the points of $C$ on $J$ are four corners of the box. These corners must then occur alternately and we may suppose them to be $X_{123}$, $X_{246}$, $X_{345}$, and $X_{156}$, as shown in the diagram. These are the four points $X_{ijk}$ such that no two suffices are congruent mod 3 and the number of suffices less than 4 is odd.

Under these circumstances any path of the form $(A, X, B, Y, A)$, where $A$ and $B$ are radiants of the box, and $X$ and $Y$ are corners not in $C$, is called an elementary re-entrant path of the *fourth* kind.

## 6.2. Homotopy

Suppose we have two paths $PR$ and $PQR$ off $C$, where $Q$ is an elementary re-entrant path of the $k$th kind. Then we call the process of deriving one of these paths from the other an *elementary deformation* of the $k$th kind. We say that two given paths $P$ and $Q$ off $C$ are *homotopic* if one can be derived from the other by a finite sequence of elementary deformations. We write this relation as $P \sim Q$, or if it is necessary to specify $C$ as $P \sim Q \bmod C$. If a path $P$, necessarily re-entrant, is homotopic to a degenerate path we say that $P$ is null-homotopic. We write this relation as $P \sim 0$, or as $P \sim 0 \bmod C$.

Homotopy is clearly an equivalence relation.

We note some simple rules of computation.

6.21 *If two path-products $PQR$ and $PQ_1R$ off $C$ are such that $Q \sim Q_1$, then $PQR \sim PQ_1R$.*

PROOF: Any sequence of elementary deformations transforming $Q$ into $Q_1$ must also transform $PQR$ into $PQ_1R$.

6.22 *If $P$ is any path off $C$ then $PP^{-1} \sim 0$.*

PROOF: If possible choose $P$ so that the theorem fails and $s(P)$ has the least value consistent with this condition.

If $s(P)=0$ the theorem is trivially true. If $s(P)=1$ it is true because $PP^{-1}$ is an elementary re-entrant path of the first kind. We deduce that $s(P) \geqslant 2$.

We can now write $P=QR$, where $s(Q)$ and $s(R)$ are both less than $s(P)$. We have $QQ^{-1} \sim 0$ and $RR^{-1} \sim 0$, by the choice of $P$. But now

$$PP^{-1} \sim QRR^{-1}Q^{-1} \sim QQ^{-1} \sim 0,$$

by 6.21. The theorem follows.

6.23 *If $PUR$ and $PVR$ are paths off $C$ such that $UV^{-1} \sim 0$, then $PUR \sim PVR$.*

PROOF: $PUR \sim PUV^{-1}(V^{-1})^{-1}R$, by 6.21 and 6.22,
$\sim PUV^{-1}VR$, by 4.321,
$\sim PVR$, by 6.21.

In what follows we prove that every re-entrant path on $M$ and off $C$ is null-homotopic,

It will be convenient to denote by $H(n)$ the proposition that every re-entrant path off $C$ which is on a flat of $M$ of dimension $\leqslant n$ is null-homotopic.

Any path $P=(X_1, X_2, \ldots, X_k)$ of $M$ determines a flat $F(P)$ satisfying

$$F(P)=X_1 \cup X_2 \cup \ldots \cup X_k.$$

6.24 *If $P$ is any path of $M$ then the flat $F(P)$ is connected.*

PROOF: $X_i \cup X_{i+1}$ is a connected line whenever $1 \leqslant i < k$. Hence the theorem follows by repeated application of 3.48.

We refer to the dimension of $F(P)$ also as the *dimension* of $P$.

## 6.3. A Lemma

We devote this section to a proof of the following proposition.

6.31 *Suppose $n \geqslant 2$ and $H(n)$ is true. Let $Q=(W, X, Y, Z, W)$ be a path off $C$ of dimension $n+1$ such that $W \cup X \cup Y$ and $Y \cup Z \cup W$ are connected planes and $W \cup Y$ is a disconnected line. Then $Q \sim 0$.*

We write $F_1=W \cup X \cup Y$ and $F_2=Y \cup Z \cup W$.

We arrive at a proof of 6.31 by way of a chain of minor propositions.

6.311 *Let $Q'=(W, X', Y, Z', W)$ be a path off $C$ such $X'$ is on $F_1$ and $Z'$ on $F_2$. Then $Q' \sim Q$.*

PROOF: This situation is illustrated in figure 6C. The proof runs as follows:

$$Q' \sim (W, X', Y)(Y, X, W)(W, X, Y)(Y, Z, W)(W, Z, Y)(Y, Z', W),$$

by 6.21 and 6.22,

$$= (W, X', Y, X, W)(W, X, Y, Z, W)(W, Z, Y, Z', W) \sim (W, X, Y, Z, W)=Q,$$

by $H(n)$, since $(W, X', Y, X, W)$ and $(W, Z, Y, Z', W)$ are on planes.

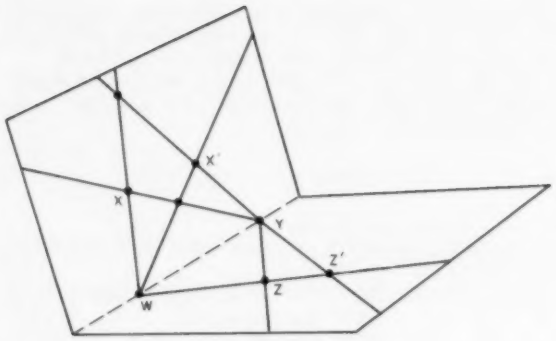



FIGURE 6C

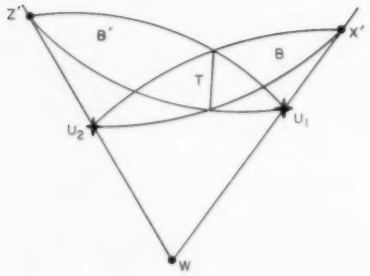



FIGURE 6D

A *transversal* of dimension $n$ is a connected $n$-flat of $M$ which is on $F(Q)$ but not on both $W$ and $Y$. Such a transversal meets each of $F_1$ and $F_2$ in a connected line, by 4.14, 4.16, and 4.281.

A *transversal* of dimension $n-1$ is a connected $(n-1)$-flat of $M$ which is on $F(Q)$ but not on $W$ or $Y$. By 4.14 and 4.16 the transversal has just one common point with each of $F_1$ and $F_2$. We call these two points the *poles* of the transversal.

Let $B$ be a transversal of dimension $n-1$ with poles $X'$ on $F_1$ and $Z'$ on $F_2$. Using 4.26 we find that $B$ is on two distinct connected $n$-flats on $F(Q)$. By 4.14 and 4.16 these are $B \cup W$ and $B \cup Y$, and each of them is a transversal of dimension $n$. Their connected lines of intersection with $F_1$ and $F_2$ are $X' \cup W$, $X' \cup Y$, $Z' \cup W$ and $Z' \cup Y$.

6.312 *Let $B$ be a transversal of dimension $n$. Then either $Q \sim 0$ or $B$ has a pole in $C$.*

PROOF: Let the poles of $B$ be $X'$ on $F_1$ and $Z'$ on $F_2$. Suppose neither of them is in $C$.

By 4.34 there is a path $R$ from $X'$ to $Z'$ on $B$ which is off $C$. Hence there are paths $(W, X')\,R(Z', W)$ and $(X', Y, Z')R^{-1}$ on the transversals $B \cup W$ and $B \cup Y$ respectively of dimension $n$. We now have

$$\begin{aligned} Q &\sim (W, X', Y, X', W), \text{ by 6.311} \\ &= (W, X')\,(X', Y, Z')\,(Z', W) \\ &\sim (W, Z')R^{-1}R(Z', W), \text{ by 6.23 and } H(n), \\ &\sim (W, Z', W) \text{ by 6.22} \\ &\sim 0, \end{aligned}$$

by an elementary deformation of the first kind. This proves the proposition.

Let us make the assumption that $Q$ is not null-homotopic.

By 4.27 there is a transversal $A$ of dimension $n$ which is not on $Y$. Let its lines of intersection with $F_1$ and $F_2$ be $L_1$ and $L_2$ respectively. By the definition of a linear subclass there is a point $X'$, other than $W$, on $L_1$ and not in $C$. By 4.27 there is a connected $(n-1)$ flat $B$ on $M$ which is on $A$ and $X'$ but not on $W$. Now $B$ is a transversal of dimension $n-1$. Let its pole on $L_2$ be $U_2$. Then $U_2 \epsilon C$, by 6.312.

Similarly there is a transversal $B'$ of dimension $n-1$ on $A$ having a point $Z'$ not in $C$ as its pole on $L_2$ and a point $U_1$ of $C$ as its pole on $L_1$. We write $T = (B \cap B')$. By 4.14 and 4.16 $T$ is an $(n-2)$-flat on $A$. (See fig. 6D.)

Let $\Sigma$ be the class of all points on $T$ which are not in $C$.

6.313 *$\Sigma$ is non-null.*

PROOF: Since $T$ is an $(n-2)$ flat on an $(n-1)$-flat $B$, and since $B$ has a point $U_2$ in $C$ and a point $X'$ not in $C$, this follows from 4.36.

6.314 *Let $T_i$ be any point of $\Sigma$. Then $T_i \cup W$ and $T_i \cup Y$ are disconnected lines.*

PROOF: Suppose the flat $T_i \cup Y$ is connected (fig. 6E). There is a path $R_0$ from $Y$ to $T_i$ on $T_i \cup Y$, by 4.34, which is off $C$. Similarly there is a path $R_1$ from $X'$ to $T_i$ on $B$ and a path $R_2$ from $Z'$ to $T_i$ on $B'$, both paths being off $C$.

Now $(X', Y)R_0R_1^{-1}$ is a re-entrant path on the transversal $B \cup Y$ of dimension $n$, and $(Y, Z')R_2R_0^{-1}$ is a re-entrant path on the transversal $B' \cup Y$ of dimension $n$. Both these paths are null-homotopic, by $H(n)$. We thus have

$$\begin{aligned} Q &\sim (W, X', Y, Z', W), \text{ by 6.311,} \\ &= (W, X')(X', Y)(Y, Z')(Z', W) \\ &\sim (W, X')R_1R_0^{-1}R_0R_2^{-1}(Z', W), \text{ by 6.23,} \\ &\sim (W, X')R_1R_2^{-1}(Z', W), \text{ by 6.22.} \end{aligned}$$

But the last path is on the $n$-flat $A$. Hence $Q \sim 0$ by $H(n)$, contrary to assumption.

We deduce that $Y \cup T_i$ has nontrivial separators $Y$ and $T_i$. Hence $Y$ and $T_i$ are the only points on it and the flat must be a disconnected line. (3.42 and 3.54.)

The transversal $B \cup Y$ of dimension $n$ is not on $W$. We can therefore repeat the construction for $B'$, using $B \cup Y$ instead of $A$ and interchanging the roles of $Y$ and $W$. There results a transversal $B''$ of dimension $n-1$ on $B \cup Y$. Its pole on the line $U_2 \cup Y$ is not in $C$ and its pole on the line $X' \cup Y$ is in $C$. It meets $B$ in an $(n-2)$-flat $T''$ analogous with $T$.

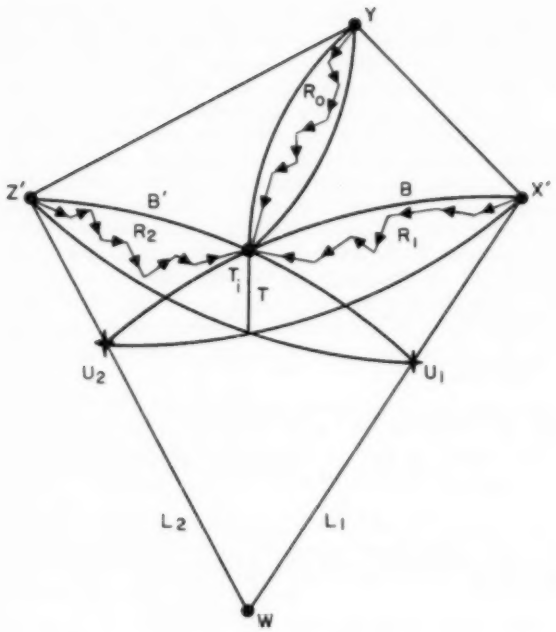



FIGURE 6E

In view of the foregoing argument we may assert that any point $T_j'$ on $T'$ which is not in $C$ has the property that $W \cup T_j'$ is a disconnected line. In particular $B''$ has a point in common with the disconnected line $Y \cup T_i$, and this point can only be $T_i$. Hence $T_i$ is on $T'$ and therefore $W \cup T_i$ is a disconnected line.

Let $K$ be a connected flat of $M$ which is on $F_1$ and $F(Q)$, and also on some point of $\Sigma$, and which has the least dimension consistent with these properties. Clearly either $F(Q)$ or one of its subsets satisfies this definition.

6.315 $dK = 3$.

PROOF: We have

$$n+1 = dF(Q) \geqslant dK > dF_1 = 2,$$

and so $dK \geqslant 3$. But $dT = n-2$. Hence, by 4.16,

$$d\langle K \cap T \rangle \geqslant dK - 3.$$

Choose a point $N$ on $\langle K \cap T \rangle$, taking $N$ in $C$ if this is possible. By 4.27 there is a connected $(dK-1)$-flat $K'$ on $F_1$ and $K$, but not on $N$. By 4.14 and 4.16 $\langle K' \cap T \rangle$ is a flat on $\langle K \cap T \rangle$ of dimension $d\langle K \cap T \rangle - 1$.

All the points of $M$ on $\langle K' \cap T \rangle$ belong to $C$, by the definition of $K$. By the choice of $N$ this implies that either $d\langle K' \cap T \rangle = -1$ or $N$ is in $C$. But in the latter case all the points of $\langle K \cap T \rangle$ are in $C$, by 4.36, which is contrary to the definition of $K$.

We deduce that $d\langle K' \cap T \rangle = -1$ and therefore $d\langle K \cap T \rangle = 0$. But we have seen that $d\langle K \cap T \rangle \geqslant dK - 3$. The proposition follows.

6.316. $n = 2$.

PROOF: Suppose $n \geqslant 3$. Then $F_2$ is not on $K$, since $dF(Q) = d(F_1 \cup F_2) \geqslant 4$, by 6.315. By 4.27 there is a connected $n$-flat $K''$ on $F_2$ and $F(Q)$ but not on $T_i$. Write $F_3 = \langle K'' \cap K \rangle$. Then $F_3$ is a plane on $K$ and $W \cup Y$, by 4.14 and 4.16.

By 4.27 there is a connected line $L$ on $K$ and $T_i$ which has no common point with the disconnected line $W \cup Y$. It meets $F_1$ and $F_3$ in distinct points $X_1$ and $X_3$, respectively, neither of which is $T_i$. (Fig. 6F.)

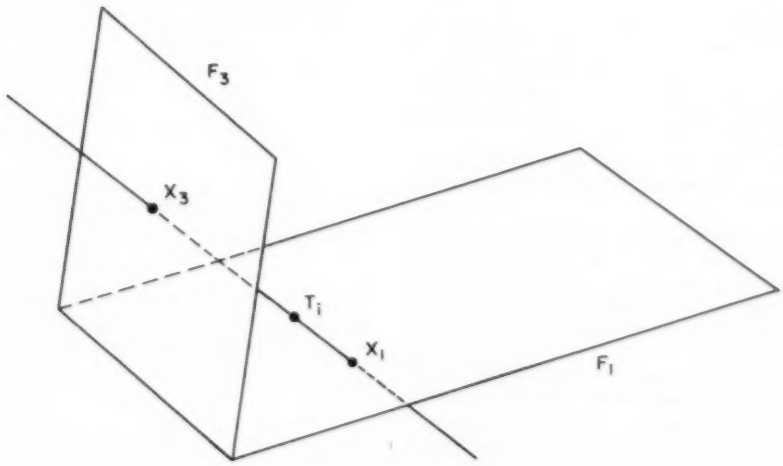



FIGURE 6F

We note that $F_3 = W \cup Y \cup X_3$, by 4.14. Moreover $X_1 \cup W$ and $X_1 \cup Y$ are connected lines, by 4.281. Hence $X_1 \cap W$ and $X_1 \cap Y$ are both non-null, by 4.22. On the other hand $T_i \cup W$ and $T_i \cup Y$ are disconnected lines, by 6.314. Hence $T_i \cap W$ and $T_i \cap Y$ are null, by 4.22. But $L = X_1 \cup T_i = X_3 \cup T_i$, by 4.22. Hence $X_3 \cap W$ and $X_3 \cap Y$ are both non-null and therefore $F_3$ is connected by 3.48.

By 4.35 there is a path $R$ from $Y$ to $W$ on $F_3$ which is off $C$. The re-entrant paths $(W, X, Y)\, R$ and $(Y, Z, W)\, R^{-1}$ are on $K$ and $K^n$ respectively. They are thus null-homotopic, by $H(n)$. Hence

$$Q = (W, X, Y)(Y, Z, W) \sim R^{-1} R \sim 0,$$

by 6.22 and 6.23.

Since $n \geqslant 2$ by the hypothesis of 6.31 we deduce, from this contradiction to our main assumption that $Q$ is not null-homotopic, that $n = 2$.

It follows from this result that $dT = 0$. Hence the flat $T$ is a point of $M$, identical with $T_i$. The three flats $W \cup Y$, $Y \cup T$ and $T \cup W$ are disconnected lines, by 6.314. Hence $W \cup Y \cup T$ is not a connected plane, by 4.282.

Any plane on the 3-flat $F(Q)$ has a common point with each of the disconnected lines $W \cup Y$, $Y \cup T$ and $T \cup W$. It is therefore on one of these lines, by 4.23. Each line of $F(Q)$ is on a plane of $F(Q)$, by 4.14. Hence each such line is on $W$, $Y$, or $T$, by 4.171.

6.317 *There are just four connected lines on each of the planes $F_1$ and $F_2$, two being on $W$ and two on $Y$. Moreover each such line is on just three points, of which just one is in $C$.*

PROOF: Suppose one such line $L_1$ is on $k$ distinct points $X_1, X_2, \ldots, X_k$ other than $Y$ or $W$. Then $k > 2$, by 4.23. Of course $L_1$ is on $Y$ or $W$.

Suppose $L_1$ is on $F_1$. By 4.27 there is a transversal $F$ of dimension 2 on $L_1$ and a transversal $B_i$ of dimension 1 on $F$ and $X_i$ for each $i$. $(i = 1, 2, 3)$. Then $F$ meets $F_2$ in a connected line $L_2$. The line $B_i$ is on $T$, since it is not on $W$ or $Y$. Hence $B_i = X_i \cup T$, and $B_i$ is uniquely determined for each $i$. Let $X_i'$ be the common point of $B_i$ and $L_2$. Since $B_1 = X_1' \cup T$ the $k$ points $X_1', X_2', \ldots, X_k$ are all distinct. But at most one point on each of the lines $L_1$ and $L_2$ is in $C$, and each $B_i$ has at least one pole in $C$. Similar reasoning applies if $L_1$ is on $F_2$.

We deduce that each connected line on $F_1$ or $F_2$ is on just three points, and that just one of these points is in $C$.

There are two connected lines on $Y$ and two on $W$ in the plane $F_1$, by 4.26. Since the lines through $W$ (or $Y$) must meet those through $Y$ (or $W$) in distinct points it follows from the preceding result that there are just two such lines on $Y$ and just two on $W$. The same result holds for $F_2$.

6.318 *Each of the disconnected lines $W \cup Y$, $Y \cup T$ and $T \cap W$ is on just two connected planes of $F(Q)$.*

PROOF: Let $L'$ and $L''$ be the two connected lines on $W$ and $F_1$. (6.317). Each is on a transversal of dimension 2, by 4.27, and both these transversals are connected planes on $T \cup W$. There is no other connected plane on $T \cup W$ since each such plane would meet $F_1$ in a connected line through $W$. (We have shown that $T \cup W \cup Y$ is not a connected plane.) A similar argument applies for $T \cup Y$.

Now suppose there is a connected plane $F_3$, distinct from $F_1$ and $F_2$, on $W \cup Y$. We note that $F_3$ is not on $T$, since $W \cup Y \cup T$ is not a connected plane.

Let $U_1$ be a point of $C$ on $F_1$. By 4.27 it is on a transversal $B$ of dimension 1. This must be on $T$. Let it meet $F_1, F_2$, and $F_3$ in $U_1, U_2$, and $U_3$ respectively. Similarly let $V_2$ be a point of $C$ on $F_2$. It is on a transversal $B'$ of dimension 1. $B'$ is on $T$ and meets $F_1, F_2$, and $F_3$ in distinct points $V_1, V_2$, and $V_3$ respectively. Since $T$ is not in $C$ we deduce that the points $U_2, U_3, V_1$, and $V_3$ are not in $C$. (Fig. 6G.)

We have

$$\begin{aligned}(W, V_1, Y) &\sim (W, V_1, V_3, W)(W, V_3, V_1, Y) \\ &\sim (W, V_1, V_3, W)(W, V_3, Y)(Y, V_3, V_1, Y) \\ &\sim (W, V_3, Y),\end{aligned}$$

by 6.21 and 6.22, with the help of $H(n) = H(2)$. Similarly

$$(W, U_2, Y) \sim (W, U_3, Y).$$

Hence

$Q \sim (W, V_1, Y, U_2, W)$, by 6.311, $\sim (W, V_3, Y, U_3, W)$, by the preceding results,

$\sim 0$, by $H(2)$.

But this is contrary to assumption. The proposition follows.

We observe that our figure on $F(Q)$ is a box with radiants $Y$, $W$, and $T$. We therefore revert to the notation of section 6.1. We put $Y = Z_1$, $W = Z_2$ and $T = Z_3$. We also identify $P_3$ with $F_1$ and $P_6$ with

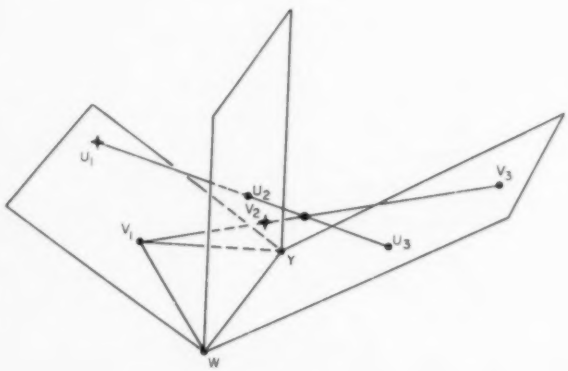



FIGURE 6G

$F_2$. Each connected line is on $Y$, $W$, or $T$, and is thus on at most one point of $C$. But each connected line on $P_3$ or $P_6$ is on one point of $C$, by 6.317. We may therefore suppose the points of C to be as shown in figure 6B.

This implies that $Q$ is an elementary re-entrant path of the fourth kind. Thus $Q \sim 0$, contrary to assumption. This contradiction establishes the lemma 6.31.

### 6.4. Homotopy Theorem

We now turn to the proof of the main theorem of the chapter.

6.41 *Every re-entrant path of M which is off C is null-homotopic*

PROOF: We proceed by induction. We observe first that $H(0)$ is trivially true. We assume as an inductive hypothesis that $H(n)$ is true for some non-negative integer $n$, and we try to deduce the truth of $H(n+1)$.

Let $P$ be any path of $M$ off $C$ which is re-entrant and has dimension $n+1$. Choose a connected $n$-flat $J$ of $M$ which is on $F(P)$ and the origin $X_0$ of $P$, as is possible by 4.25 and 6.24.

Let $R$ be any re-entrant path on $F(P)$ with origin $X_0$. Write

$$R=(X_0, X_1, X_2, \ldots, X_m, X_0).$$

We write $u(R)$ for the number of terms of $R$, counting repetitions, which are not on $J$. If $u(R)>0$ we denote by $X_i$ the first term of $R$ which is not on $J$. We then write $v(R)$ for the dimension of the connected flat $X_{i-1}\cup X_i\cup X_{i+1}$, taking $X_{m+1}=X_0$ if $i=m$. If $u(R)=0$ we write $v(R)=0$.

We may suppose $R$ chosen so as to satisfy the following conditions

(i) $R \sim P$

(ii) $u(R)$ has the least value consistent with (i),

(iii) $v(R)$ has the least value consistent with (i) and (ii).

We write $F=X_{i-1}\cup X_i\cup X_{i+1}$. We also write $R$ in the form

$$R_1(X_{i-1}, X_i, X_{i+1})R_2$$

so that $R_1$ is a path on $J$.

We assume $u(R)>0$. This implies $v(R)>0$. We consider separately the cases $v(R)=1$, $v(R)=2$ and $v(R)\geq 3$.

*Case I* $v(R)=1$.

In this case $F$ is a connected line. If $X_{i+1}=X_{i-1}$ we have $P \sim R \sim R_1R_2$, by an elementary deformation of the first kind. Since $u(R_1R_2)<u(R)$ this is contrary to the choice of $R$.

If $X_{i+1}\neq X_{i-1}$, then

$$(X_{i-1}, X_i, X_{i+1}, X_{i-1})$$

is an elementary re-entrant path of the second kind.

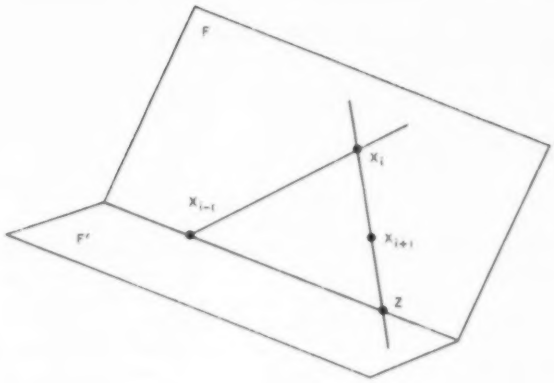



FIGURE 6H

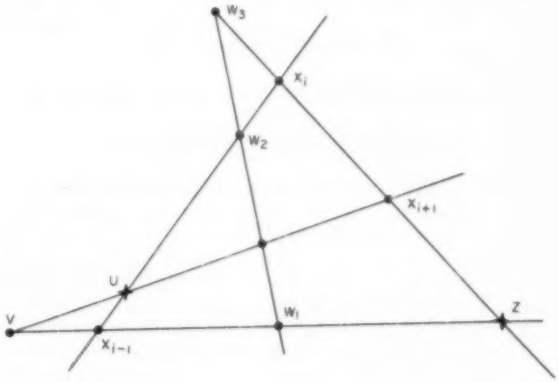



FIGURE 6I

Hence

$$P \sim R \sim R_1(X_{i-1}, X_{i+1})R_2)$$

by 6.23. But this is impossible since

$$u(R_1(X_{i-1}, X_{i+1})R_2) < u(R).$$

*Case II:* $v(R)=2$.

$F$ is now a connected plane on $F(P)$. It meets $J$ in a line $L$, by 4.14 and 4.16. Let $Z$ be the point of intersection of the lines $L$ and $X_i\cup X_{i+1}$ on $F$. (Figs. 6H and 6I.)

Suppose $Z$ is not in $C$, as in figure 6H. We define $Q$ as the degenerate path $(Z)$ if $Z=X_{i+1}$, and as the path $(Z, X_{i+1})$ otherwise. Then

$$(X_i, X_{i+1})\ Q^{-1}\ (Z, X_i)$$

is an elementary re-entrant path of the first or second kind. Hence

$$R \sim R_1(X_{i-1}, X_i, Z)QR_2,$$

by 6.23. If $L$ is connected we have

$$(X_{i-1}, X_i, Z, X_{i-1}) \sim 0,$$

by an elementary deformation of the second kind, and so

$$R \sim R_1(X_{i-1}, Z)\ QR_2,$$

by 6.23.

If $L$ is not connected it is on a connected plane $F'$ contained in $J$, by 4.28. We can find a connected line $L'$ on $X_{i-1}$ and $F'$ by 4.26, and a point $T$ on $L'$ distinct from $X_{i-1}$ and not in $C$, by 4.23. The flats $T \cup X_{i-1}$ and $T \cup Z$ are connected lines, by 4.282. Hence

$$(X_{i-1}, X_i, Z, T, X_{i-1}) \sim 0,$$

by 6.31. So by 6.23 we have

$$R \sim R_1(X_{i-1}, T, Z)\ QR_2.$$

Thus, whether or not $L$ is connected, we have

$$R \sim R_3QR_2,$$

where $R_3$ is on $J$. But this is contrary to the choice of $R$, since $u(R_3QR_2) < u(R)$.

We may now suppose $Z \epsilon C$, as shown in figure 6I. In this case there is a connected line $L'$ other than $X_i \cup X_{i+1}$ on $X_{i+1}$ and $F$, by 4.26. If $L'$ is on $X_{i-1}$ then $(X_{i-1}, X_i, X_{i+1}, X_{i-1})$ is an elementary re-entrant path of the second kind. We therefore have

$$R \sim R_1(X_{i-1}, X_{i+1})R_2, = R' \text{ say},$$

by 6.23. This is contrary to the choice of $R$, since $u(R') < u(R)$. We deduce that in fact $L'$ is not on $X_{i-1}$. It therefore meets the lines $X_{i-1} \cup X_i$ and $L$ in distinct points $U$ and $V$ respectively. Since $Z$ is in $C$ and $X_{i-1}$ is not, the point $V$ is not in $C$.

Suppose $U$ is not in $C$, contrary to the indication given in figure 6I. Using 6.23 with elementary re-entrant paths of the second kind we have

$$\begin{aligned} R &\sim R_1(X_{i-1}, U, X_i, X_{i+1})R_2 \\ &\sim R_1(X_{i-1}, V, U, X_{i+1})R_2 \\ &\sim R_1(X_{i-1}, V, X_{i+1})R_2, \qquad = R' \text{ say}. \end{aligned}$$

But then $u(R') < u(R)$, contrary to the choice of $R$.

We deduce that $U$ is in $C$, as indicated in figure 6I. In this case it may happen that each connected line on $F$ is on either $U$ or $Z$. Then $(X_{i-1}, X_i, X_{i+1}, V, X_{i-1})$ is an elementary re-entrant path of the third kind. Using 6.23 we deduce that

$$R \sim R(X_{i-1}, V, X_{i+1})R_2,$$

which is impossible, as before.

It follows that there is a connected line $L''$ on $F$ which is not on $U$ or $Z$. Let it cut the lines $L, X_{i-1} \cup X_i$ and $X_i \cup X_{i+1}$ in the points $W_1$, $W_2$, and $W_3$ respectively. These points are distinct from $U$ and $Z$ and are therefore not in $C$.

If $L''$ is on $X_{i+1}$ we can substitute it for $L'$ in the preceding argument. This reduces the problem to the case in which $U$ is not in $C$, and so yields a contradiction. We may accordingly suppose that $L''$ is not on $X_{i+1}$.

If $L''$ is on $X_i$ it meets $L$ in a point $W_1$ distinct from $X_{i-1}$ and $Z$.
Writing

$$R' = R_1(X_{i-1}, W_1, X_i, X_{i+1})R_2,$$

we then have $R \sim R'$, by an elementary deformation of the second kind. If $L''$ is not on $X_i$ then the points $W_2$, $W_3$, $X_i$, $X_{i+1}$, and $Z$ are all distinct. If $L''$ is then on $X_{i-1}$ we write

$$R' = R_1(X_{i-1}, W_3, X_{i+1})R_2,$$

and have $R \sim R'$, by elementary deformations. If however $L''$ is not on $X_{i-1}$ the points $X_{i-1}$ and $W_2$ are distinct. We then write

$$R' = R_1(X_{i-1}, W_1, W_3, X_{i+1})R_2$$

and have

$$\begin{aligned} R &\sim R_1(X_{i-1}, W_2, X_i, X_{i+1})R_2 \\ &\sim R_1(X_{i-1}, W_1, W_2, W_3, X_i, X_{i+1})R_2 \\ &\sim R', \end{aligned}$$

by 6.23. For each of these three possibilities we have

$$R \sim R',\ u(R') = u(R),\ v(R') = v(R) = 2.$$

Hence we may replace $R$ by $R'$ in the preceding argument. This reduces the problem to the case in which $U$ is not in $C$, and so yields a contradiction.

This completes the analysis of Case II.

Case III. $v(R) \geqslant 3$.

There is a connected plane $K$ on $F$ and the line $X_{i-1} \cup X_1$. This plane meets $J$ in a line $L$. Choose a point $N$, distinct from $X_{i-1}$, on $L$ and if possible in $C$. By 4.27 there is a connected $(v(R)-1)$-flat $F'$ on $X_i \cup X_{i+1}$ and $F$, but not on $N$. Now $F'$ is not on $X_{i-1}$, for otherwise we would have $F = X_{i-1} \cup X_i \cup X_{i+1} \subseteq F'$. Hence $F'$ meets $L$ in a point $N'$ distinct from $N$ and $X_{i-1}$. It follows that $L$ is connected, by 4.23, and that $N'$ is not in $C$. (See fig. 6J.)

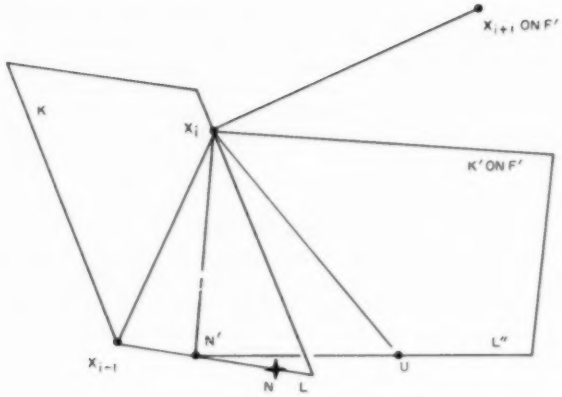



FIGURE 6J

The flats $F'$ and $K$ intersect in a line $L'$ on $X_i$ and $N'$. If $L'$ is connected we write

$$R' = R_1(X_{i-1}, N', X_i, X_{i+1})R_2,$$

and have $R' \sim R$ by 6.23. If $L'$ is not connected it is on a connected plane $K'$ on $F'$, by 4.28. This meets $J$ in a connected line $L''$ on $N'$, by 4.282. We can find a point $U$ on $L''$ distinct from $N'$ and not in $C$. The flat $U \cup X_i$ is a connected line, by 4.282. We now have

$$(N', U, X_i, X_{i-1}, N') \sim 0,$$

by 6.31. In this case we write

$$R' = R_1(X_{i-1}, N', U, X_i, X_{i+1})R_2.$$

We then have

$$\begin{aligned} R' &\sim R_1(X_{i-1}, N', U, X_i, X_{i-1}, N', X_{i-1}, X_i, X_{i+1})R_2 \\ &\sim R_1(X_{i-1}, N', X_{i-1}, X_i, X_{i+1})R_2 \\ &\sim R_1(X_{i-1}, X_i, X_{i+1})R_2 \sim R, \end{aligned}$$

by 6.23 and an elementary deformation.

So, whether $L'$ is connected or not, we have

$$R' \sim R, \qquad u(R') = u(R), \qquad v(R') < v(R).$$

But this is contrary to the definition of $R$.

From the above analysis of the hypothesis $u(R) > 0$ we deduce that in fact $u(R) = 0$. Hence $R$ is on $J$ and has dimension $\leq n$, and therefore $R \sim 0$, by $H(n)$. Hence $P \sim 0$.

We have now shown that $H(n+1)$ is deducible from $H(n)$. Hence the proposition $H(n)$ is true for every non-negative integer $n$ (since $H(0)$ is true), by induction. Thus our theorem holds.

### 6.5. Special Cases

When $C$ is null there are only two kinds of elementary re-entrant path to consider, the first and the second. The homotopy theorem then tells us that every re-entrant path in a matroid can be deformed into a degenerate one by a sequence of operations each of which is confined to a single line or plane.

The homotopy theorem can of course be applied to the polygon-matroid of a graph $G$. But care must be taken to distinguish between the paths of $P(G)$ and those of $G$ itself. A path in $P(G)$ corresponds to a sequence of polygons in $G$ such that the union of two consecutive polygons is a nonseparable subgraph $H$ such that $P(H)$ has rank 2. It can be shown that these conditions on $H$ imply that $H$ is a "$\theta$-graph", made up of three arcs such that any two have both ends but no other edge or vertex in common.

The homotopy theorem with $C$ null asserts that a re-entrant sequence of polygons of the above kind can be reduced to a degenerate sequence by operations each of which is confined to a single polygon or $\theta$-graph of $G$.

## 7. Characterization of Regular Matroids

### 7.1. Some Preliminary Observations

In this chapter we resume the discussion of regular chain-groups and matrioids which was broken off at the end of section 5.5. Our task is to show that a binary matroid is regular if it has no minor of Type BI or BII.

We first note some simple theorems which will be helpful in the discussion.

7.11 *Let $S$ be an $(n-1)$-flat on an $n$-flat $T$ of a matroid $M$. Let $a$ be any cell of $T-S$. Then $\langle T-\{a\}\rangle = S$.*

This follows from 4.13 and 4.14.

7.12 *Let $X$, $Y$ and $Z$ be distinct collinear points of a binary matroid $M$. Then $X+Y+Z$ is null.*

A proof of this theorem is contained in the proof of 5.35.

7.13 *Let $S$ be an $n$-flat on an $(n+2)$-flat $T$ of a binary matroid $M$, where $n \geq 0$. Then there are at most three $(n+1)$-flats of $M$ which are on both $S$ and $T$.*

PROOF: Suppose there are $k$ such $(n+1)$-flats.

By 4.27 there is a line $L$ on $T$ which has no common point with $S$. The $k(n+1)$-flats meet $L$ in $k$ distinct points. Hence $k \leq 3$, by 5.35 and 4.23.

### 7.2. Fano Configuration

We can define a *Fano configuration* in a binary matroid $M$ as a figure made up of a plane $P$ of $M$, seven distinct lines on $P$, and the points of intersection of these seven lines.

Suppose $F$ is such a configuration. Any line $L$ of $F$ is on at most three points, by 5.35, and each of these is on at most two other lines on $P$. But each line on $P$ must have a common point with $L$. We deduce that each line of $F$ is on three distinct points of $F$, each of which is on exactly two other lines of $F$. Hence $F$

has just seven points. Moreover there can be no eighth line on $P$, and therefore there is no eighth point on $P$, by 4.25 and 5.35.

We enumerate the seven lines as $L_1, L_2, \ldots, L_7$. We denote the point of intersection, if any, of three lines $L_i$, $L_j$, and $L_k$ by $X_{ijk}$.

We may assume a point $X_{135}$ on $L_1$, $L_3$, and $L_5$. The line $L_7$, which is not on $X_{135}$, meets $L_1$, $L_3$, and $L_5$ in points we may label $X_{147}$, $X_{367}$, and $X_{257}$ respectively. The lines $L_2$ and $L_4$ each meet $L_3$ in its point other than $X_{135}$ and $X_{367}$, which point is therefore $X_{234}$. Similarly the remaining points of the figure are $X_{456}$ and $X_{126}$.

A Fano configuration is shown in figure 7A, in which one of the seven "lines" is indicated by a curve.

The seven lines are connected, by 4.23, and so $P$ is connected, by 3.48.

7.21 *If a binary matroid $M$ includes a Fano configuration, then it has a minor of Type BI.*

PROOF: Let $P$ be the plane of the Fano configuration. There is a binary chain-group $N$ such that $M(N) = M \times P$, by 5.31.

Consider the chains $f$, $g$, and $h$ on $P$ over $GF(2)$ whose domains are $X_{126}$, $X_{234}$, and $X_{456}$ respectively. They are linearly independent chains of $N$ since $X_{126}$ is not on $L_4$, that is not contained in $X_{234} \cup X_{456}$. Moreover they generate the chains corresponding to all the atoms of $M \times P$, by 7.12. Hence their representative vectors can be taken as the three rows of a representative matrix $K$ of $N$, by 5.32.

Since $P$ is a union of atoms of $M$ the matrix $K$ has no zero column. If it has two equal columns corresponding to cells $a$ and $b$ it is clear that an atom of $M \times P$ contains $a$ if and only if it contains $b$. But then $\langle P - \{a\} \rangle = \langle P - \{b\} \rangle$.

Since there are seven distinct lines on $P$ it follows from 7.11 that all the seven distinct nonzero 3-vectors over $GF(2)$ occur as columns of $K$. Hence, there exists $S \subseteq P$ such that $K(S)$ is the first matrix of section 5.5, apart from a permutation of columns. But then $(M \times P) \cdot S$ is a matroid of Type BI, by 5.21. Since $(M \times P) \cdot S$ is a minor of $M$ the theorem follows.

### 7.3. Heptahedron

We use the term "heptahedron" to denote a figure on a 3-space $F$ of a binary matroid $M$, defined by seven distinct planes such that no three are on a common line. We include in the figure the 21 lines of intersection of the seven planes, and those points of $F$ which are common to three or more of the seven planes.

We enumerate the seven planes as $P_1, P_2, \ldots, P_7$, and write $L_{ij}$ for the line of intersection of $P_i$ and $P_j$, taking $0 \leq i < j \leq 7$.

A point on $L_{ij}$ is on at most three lines of $P_i$, by 7.13. On the other hand there are exactly six lines of the heptahedron on $P_i$. It follows that each of the lines $L_{ij}$ is on exactly three points of the figure, that one of these points is on exactly three of the seven planes, and that each of the two other points is on exactly four of them (5.35). Hence each of the lines $L_{ij}$ is connected. It follows that the planes $P_i$ and the flat $F$ are connected.

We write $X_{ijk}$ for a point which is on just three planes $P_i$, $P_j$, and $P_k$ of the heptahedron, and $X_{ijkl}$ for a point on just four such planes $P_i$, $P_j$, $P_k$, and $P_l$. Each point of the first kind is on exactly three lines of the heptahedron, and each point of the second kind on exactly six. It follows that the figure has just seven points of each kind.

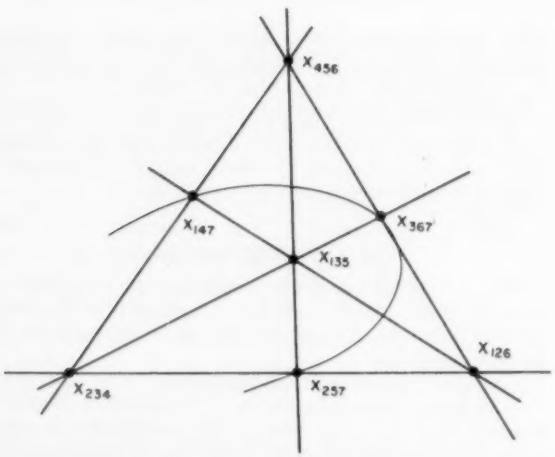



FIGURE 7A

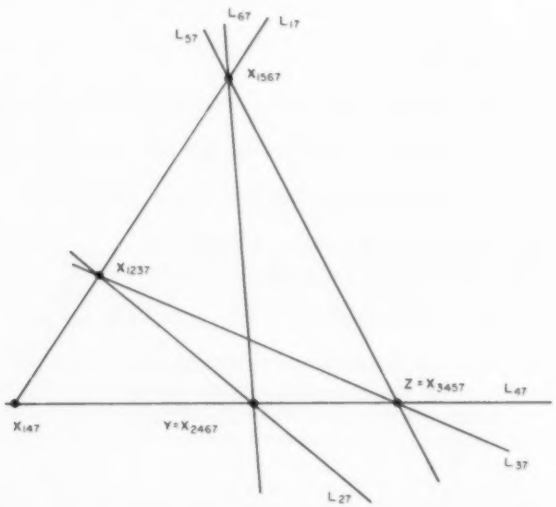



FIGURE 7B

We may assume $L_{17}$ to be on $X_{147}$, $X_{1237}$, and $X_{1567}$. The line $L_{47}$ is on two points $Y$ and $Z$ distinct from $X_{147}$. On $P_7$ each of these is on three lines of the heptahedron meeting $L_{17}$ in three distinct points. We may therefore suppose $Y$ to be on $L_{27}$ and $L_{67}$ while $Z$ is on $L_{37}$ and $L_{57}$. Then $Y=X_{2467}$ and $Z=X_{3457}$. (Fig. 7B.)

The remaining points of the heptahedron on $P_7$ are $X_{367}=\langle L_{37}\cap L_{67}\rangle$ and $X_{257}=\langle L_{27}\cap L_{57}\rangle$.

The lines $L_{13}$ and $L_{14}$ on $P_1$ meet in a point $X_{134x}$, since $L_{14}$ is on $X_{147}$. Since no three planes of the heptahedron are on a common line this point cannot have three suffices in common with any of the points on $P_7$. It is therefore $X_{1346}$. But $X_{1346}$ is on the lines $L_{13}$, $L_{14}$, $L_{16}$, $L_{34}$, $L_{36}$, and $L_{46}$. These meet $P_7$ in $X_{1237}$, $X_{147}$, $X_{1567}$, $X_{3457}$, $X_{367}$, and $X_{2467}$ respectively. The third points on these lines are therefore $X_{135}$, $X_{1245}$, $X_{126}$, $X_{234}$, $X_{2356}$, and $X_{456}$.

We have now accounted for all the points of the figure. Those of the first kind are

$$X_{126},\ X_{135},\ X_{147},\ X_{234},\ X_{257},\ X_{367},\ X_{456},$$

and those of the second kind are

$$X_{1237},\ X_{1245},\ X_{1346},\ X_{1567},\ X_{2356},\ X_{2467},\ X_{3457}.$$

A heptahedron can be constructed from the box of figure 6B. To do this we replace the three disconnected lines by connected ones, and we imagine a seventh plane to pass through the points of $C$ and to cut the lines $Z_iZ_j$ in their third points. Since any two points of $C$ determine a unique plane of the box it is clear that no line of the resulting figure is on three of the seven planes. The notation of figure 6B is not that used above.

7.31 *If a binary matroid $M$ includes a heptahedron, then it has a minor of Type BII.*

PROOF: We use for the heptahedron the notation set out above. There is a binary chain-group $N$ such that $M(N)=M\times F$, by 5.31. The rank of $M(N)$ is 4.

Let $f$, $g$, $h$, and $j$ be the chains on $F$ over $GF(2)$ whose domains are $X_{234}$, $X_{1346}$, $X_{1245}$, and $X_{1237}$ respectively. Any three of these atoms define a plane which is not on the fourth, so the four chains are linearly independent. Using 7.12 we can verify that they generate the chains corresponding to the atoms of $M(N)$, and therefore all the chains of $N$ by 5.32. Hence their representative vectors can be taken as the four rows of a representative matrix $K$ of $N$.

For each $P_i$ we select a cell $a_i\epsilon F-P_i$. We then have $P_i=\langle F-\{a_i\}\rangle$, by 7.11. We write $S=\{a_1, a_2, \ldots, a_7\}$. We further suppose that the enumeration of the cells of $F$, used in defining $K$, is such that the columns representing the cells $a_i$ occur in the order of their suffices.

Now a plane $P_i$ passes through a point $X$ of the heptahedron if and only if $i$ is one of the suffices of $X$. Moreover this happens if and only if $a_i$ is not a cell of $X$. We can therefore evaluate the matrix $K(S)$ as

$$\begin{array}{c} a_1\ a_2\ a_3\ a_4\ a_5\ a_6\ a_7 \\ \left(\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right) \end{array}$$

This is a representative matrix of the minor $M_1=(M\times F)\cdot S$ of $M$, by 5.21. But then $M_1{}^*$ is a matroid of Type BI, by 5.23. Thus $M_1$ is of Type BII, and the theorem follows.

Theorems 7.21 and 7.31 have simple converses. It is easily verified that a matroid of Type BI or BII has a Fano configuration or heptahedron on its plane or 3-space. If the matroid is a minor $(M\times S)\cdot T$ of a larger matroid $M$, a corresponding configuration must occur on any carrier $Z$ of $(M\times S)\cdot T$ in $M\times S$. (See section 4.4). The configuration then occurs on the flat $\langle Z\rangle$ of $M$.

## 7.4. Condition for a Regular Matroid

We devote this section to a proof of the following theorem.

7.41 *A binary matroid $M$ on a set $E$ is regular if it includes no Fano configuration and no heptahedron.*

PROOF: Assume the theorem false. Then there exists a matroid $M$, on a set $E$, which is binary and includes no Fano configuration or heptahedron, but which is not regular. Choose such a matroid $M$ so that $|E|$ has the least possible value.

Clearly $|E|\geq 1$. Choose $a\epsilon E$ and write $M'=M\cdot(E-\{a\})$. Then $M'$ has a carrier $Z$ in $M$, and $Z$ is either $E$ or $E-\{a\}$. By section 4.4 there is a $1-1$ mapping $\theta$ of the set of flats of $M'$ onto a subset of the set of flats of $M$ which preserves dimension and geometrical incidence.

As in the case of 6.31 we arrive at a proof of 7.41 through a sequence of subsidiary theorems.

7.411 *$M'$ is regular.*

PROOF: $M'$ is binary, by 5.31. It includes no Fano configuration or heptahedron, for otherwise $\theta$ would induce another such figure in $M$. Hence $M'$ is regular, by the choice of $M$.

By 7.411 there is a regular chain-group $N'$ on $E-\{a\}$ such that $M'=M(N')$. For each atom $X$ of $M'$ we select a primitive chain $g(X)$ of $N'$ such that $||g(X)||=X$. For each $b\epsilon E-\{a\}$ we denote the coefficient of $b$ in $g(X)$ by $s(X, b)$. Thus $s(X, b)$ is 1 or $-1$ if $b\epsilon X$, and is zero otherwise.

Let $C$ denote the class of all points $T$ of $M'$ such that $a\notin\theta T$.

7.412 *$C$ is a linear subclass of $M'$.*

PROOF: Suppose $X$, $Y$, and $Z$ are distinct points of $M'$ on a line $L$ of $M'$, and that $X$ and $Y$ belong to $C$.

Then $\theta X$, $\theta Y$, and $\theta Z$ are distinct points of $M$ on $\theta L$, and $\theta L$ is a line of $M$. But now $a \notin \theta X \cup \theta Y = \theta L$, by 4.22, and therefore $a \notin \theta Z$. Hence $Z \in C$. The proposition follows.

Let $X$ and $Y$ be any two distinct points of $M'$, not in $C$, which are on the same connected line $L$ of $M'$. By 4.22 there exists a cell $d \in X \cap Y$. We now write

$$t(X, Y) = s(X, d)s(Y, d).$$

7.413 *If $X$ and $Y$ satisfy the conditions just stated, then $t(X, Y)$ is uniquely determined.*

PROOF: Suppose this proposition is false for some $X$ and $Y$. Then there are distinct cells $d$ and $e$ of $X \cap Y$ such that

$$s(X, d)s(Y, d) = -s(X, e)s(Y, e)$$

Without loss of generality we may assume that

$$s(X, d) + s(Y, d) = 0,$$

$$s(X, e) + s(Y, e) = \pm 2.$$

The third point $Z$ on $L$ is the mod 2 sum of $X$ and $Y$, by Hence we have

$$Z \subset ||g(X) + g(Y)|| \subset L,$$

since $e$ is in $||g(X) + g(Y)||$ but not $Z$, and $d$ is in $L$ but not $\|g(X) + g(Y)\|$. But $\|g(X) + g(Y)\|$ is a flat of $M'$, by 5.43. Hence this result implies that $dL \geqslant 2$, by 4.14, which is contrary to the definition of a line. The proposition follows.

If $R = (X_1, \ldots, X_k)$ is any nondegenerate path off $C$ in $M'$ we write

$$u(R) = \prod_{i=1}^{k-1} t(X_i, X_{i+1})$$

7.414 *If $R$ is any nondegenerate re-entrant path off $C$ on $M'$, then $u(R) = 1$.*

PROOF: We first note that by 6.41, there is a positive integer $m(R)$ which is the least number of elementary deformations, with respect to $C$, required to convert $R$ into a degenerate path.

Assume the proposition false. Then we can find a nondegenerate re-entrant path $Q$ off $C$ on $M'$ such that $u(Q) = -1$. and $m(Q)$ has the least value consistent with this condition.

If $m(Q) > 1$ we can write either $Q = Q_1UQ_2$ and $Q' = Q_1Q_2$ or $Q' = Q_1UQ_2$ and $Q = Q_1Q_2$, where $U$ is an elementary re-entrant path with respect to $C$ and $m(Q') = m(Q) - 1 > 0$. In either case we have

$$u(Q) = u(Q) \cdot u(Q').$$

But this is impossible, since $u(Q) = u(Q') = 1$ by the choice of $Q$. We deduce that in fact $m(Q) = 1$. In other words $Q$ is an elementary re-entrant path with respect to $C$.

We must now consider four cases, one for each kind of elementary re-entrant path.

Case I. *$Q$ is of the first kind.*

We have $Q = (X, Y, X)$ for some points $X$ and $Y$. Hence

$$u(Q) = \{t(X, Y)\}^2 = 1.$$

Case II. *$Q$ is of the second kind.*

In this case we have $Q = (X, Y, Z, X)$, where $X$, $Y$, and $Z$ are distinct points on the same line or plane of $M'$.

If they are on the same line $L$ of $M'$, then $\theta X$, $\theta Y$ and $\theta Z$ are distinct points on the line $\theta L$ of $M$. But $a \in \theta X \cup \theta Y \cup \theta Z$ and therefore $\theta L$ has a fourth point $\langle \theta L - \{a\} \rangle$. This is impossible since $M$ is binary. (5.35).

We may now suppose that $X$, $Y$, and $Z$ are on the same plane $P$ of $M'$, but not on a common line. $P$ is connected, by 6.24 Let $X_1$, $Y_1$, and $Z_1$ be the third points on the connected lines $Y \cup Z$, $Z \cup X$ and $X \cup Y$ of $M'$ respectively

Suppose first that two of the points $X_1$, $Y_1$, and $Z_1$ are on a common line $L$ of $M'$. There is a cell $d$ of $P$ such that $L = \langle P - \{d\} \rangle$, by 7.11. It is clear that $L$ is not on $X$, $Y$, or $Z$. Hence $d \in X \cap Y \cap Z$. We thus have

$$u(Q) = s(X, d)s(Y, d)s(Y, d)s(Z, d)s(Z, d)s(X, d) = 1,$$

which is contrary to the choice of $Q$.

We conclude that no two of $X_1$, $Y_1$, and $Z_1$ are on a common line of $M'$. Any line through one of them is therefore on $X$, $Y$, or $Z$, by 5.35. Hence, by 4.26, there are six distinct lines on $P$, each on one of the points $X$, $Y$, and $Z$.

These six lines are mapped by $\theta$ onto six distinct lines of $M$ on the plane $\theta P$, each being on one of the points $\theta X$, $\theta Y$, and $\theta Z$ and therefore having $a$ as a cell. But then there is a seventh line $\langle \theta P - \{a\} \rangle$ on $\theta P$, by 7.11. Accordingly $M$ includes a Fano configuration.

Case III. *$Q$ is of the third kind.*

We now have $Q = (X, Y, Z, T, X)$, where $X$, $Y$, $Z$, and $T$ are distinct points on a plane $P$ of $M'$ such that no three are on a common line. We can therefore find cells $b \in P - (X \cup Y)$, $c \in P - (Y \cup Z)$, $d \in P - (Z \cup T)$, and $e \in P - (T \cup X)$. Then $b$ belongs to $Z$ and $T$ but not to $X$ or $Y$, $c$ belongs to $T$ and $X$ but not to $Y$ or $Z$, and so on.

If the chains $g(X)$, $g(Y)$, $g(Z)$, and $g(T)$ are linearly independent we can construct a nonzero linear combination of them, $h$ say, such that $h(b) = h(c) = h(d) = 0$. Then there is a point $U$ of $M'$ contained in $\|h\|$, by 5.43. This point is common to the lines $\langle P - \{b\} \rangle = X \cup Y$, $\langle P - \{c\} \rangle = Y \cup Z$ and $\langle P - \{d\} \rangle = Z \cup T$. Hence $U = Y = Z$. But this is impossible since $Y$ and $Z$ are distinct. We deduce that $g(X)$, $g(Y)$, $g(Z)$, and $g(T)$ are in fact linearly dependent. We thus have

$$\begin{vmatrix} 0 & 0 & s(T,c) & s(T,b) \\ s(Z,e) & 0 & 0 & s(Z,b) \\ s(Y,e) & s(Y,d) & 0 & 0 \\ 0 & s(X,d) & s(X,c) & 0 \end{vmatrix} = 0,$$

that is

$$s(Z,b)s(T,c)s(X,d)s(Y,e) - s(T,b)s(X,c)s(Y,d)s(Z,e) = 0$$

It follows from this result that

$$u(Q) = t(X,Y)t(Y,Z)t(Z,T)t(T,X) = 1.$$

Case IV. *Q is of the fourth kind.*

In this case there is a box on some 3-flat $J$ of $M'$ such that the six sides $P_1, P_2, \ldots, P_6$ of the box have the following properties, Each is on just two points of $C$ and no two of the $P_i$ have more than one point of $C$ in common. By the definition of a box no three of the planes $P_i$ are on a common line.

It follows that no three of the six planes $\theta P_i$ of $M$ on $\theta J$ have a common line, and that each of them has $a$ as a cell. Moreover each line of intersection of two planes $\theta P_i$ is of the form $\theta L$, where $L$ is the line of intersection of the corresponding planes $P_i$ in $M'$. $L$ is on some point of $M'$ not in $C$ and therefore $\theta L$ has $a$ as a cell.

The plane $\langle \theta J - \{a\} \rangle$ of $M$ on $\theta J$ is thus distinct from each of the six planes $\theta P_i$ and is not on any of their lines of intersection. Hence $M$ includes a heptahedron.

In each of the four cases we have deduced a contradiction, either of the hypothesis of 7.41 or of our assumption that 7.414 is false. We deduce that 7.414 is in fact valid.

We can partition the set of points of $M'$ not in $C$ into disjoint non-null classes $K_1, K_2, \ldots, K_q$ such that for any two points $X$ and $Y$ of $M'$ off $C$ there is a path from $X$ to $Y$ off $C$ in $M'$ if and only if $X$ and $Y$ belong to the same class $K_j$. For each class $K_j$ we select an arbitrary representative $T_j$ and write

$$v(T_j) = 1.$$

For any other point $X$ in $K_j$ we write

$$v(X) = u(Q)v(T_i),$$

where $Q$ is any path from $T_i$ to $X$ off $C$ in $M'$.

This definition determines the integer $v(X)$ uniquely. For suppose $Q'$ is another path from $T_i$ to $X$ off $C$ in $M'$. Then

$$u(Q) = u(Q^{-1}) = u(Q^{-1})u(Q')u(Q') \cdot = u(Q^{-1}Q')u(Q') = u(Q'),$$

by 7.414 since $Q^{-1}Q'$ is re-entrant.

For each $U \in M$ we define a chain $f_U$ over the ring of integers as follows.

Suppose first that $a \in U$. If $U = \{a\}$ we write $f_U(a) = 1$, and $f_U(b) = 0$ whenever $b \neq a$. If $U \neq \{a\}$, then $U - \{a\}$ is a non-null flat of $M'$, by 4.45. The flat $\theta(U - \{a\})$ of $M$ is $U - \{a\}$ or $U$, and the former alternative is ruled out by Axiom I. Hence $U - \{a\}$ is a point $X$ of $M'$ not in $C$. We choose $f_U$ to agree with $g(X)$ in $E - \{a\}$, and to satisfy

$$f_U(A) = v(X).$$

Suppose next that $a \notin U$. Then $U$ is a flat of $M'$, by 4.45, and $\theta U$ is $U$ or $U \cup \{a\}$.

If $\theta U = U$, then $U$ is a point of $M'$. We then choose $f_U$ to agree with $g(U)$ in $E - \{a\}$ and to satisfy

$$f_U(a) = 0$$

If instead $\theta U = U \cup \{a\}$ then $U \cup \{a\}$ is a line of $M$, by 4.13, and $U$ is a line of $M'$, by 4.44. Now each point on $U$ in $M'$ is mapped by $\theta$ onto a point of $M$ on $U \cup \{a\}$. But there are at most three points of $M$ on $U \cup \{a\}$, by 5.35, and one of these is U. Hence U must be a disconnected line of $M'$, having only two points $X$ and $Y$. Write $V = \theta X$ and $W = \theta Y$. Then $V = X \cup \{a\}$ and $W = Y \cup \{a\}$, by Axiom I. Thus $f_V$ and $f_W$ are already defined. We take either $f_U$ or $-f_U$ to be

$$f_V(a)f_W - f_W(a)f_V.$$

We note that for each atom $U$ of $M$ we have $\|f_U\| = U$ and that the coefficients in each $f_U$ are restricted to the values $1, -1$ and 0. Moreover the restriction of $f_U$ to $E - \{a\}$ is a chain of $N'$.

7.415 *Let $U$, $V$, and $W$ be the three points on a connected line $L$ of $M$. Then $f_U$, $f_V$ and $f_W$ are linearly dependent.*

PROOF: We discuss first the case $a \in L$.

We may suppose $a \in V \cup W$ and $a \notin U$, by 7.12. Then there are points $X = V - \{a\}$ and $Y = W - \{a\}$ of $M'$ such that $\theta X = V$ and $\theta Y = W$. Write $L' = L \cap (E - \{a\}) = L - \{a\}$. Then $L'$ is a flat of $M'$, by 4.45. It clearly satisfies $\theta L' = L$ and is therefore a line of $M'$. The set $U$ is also a flat of $M'$, by 4.45. It is either the line $L'$ or a point of $M'$ on it.

If $U$ is a point of $M'$ on $L'$, then $\theta U = U$, and $U$ is distinct from $X$ and $Y$. The line $L'$ is thus connected, by 4.23. We can define $v(X)$ by a path $Q$ off $C$ from a representative point $T_i$ to $X$, and then define $v(Y)$ by the path $Q(X, Y)$. We then have

$$v(X)v(Y) = u(Q)v(T_i) \cdot u(Q)t(X,Y)v(T_i) = t(X,Y).$$

Hence for each $b \in X \cap Y$ we have $v(X)v(Y) = s(X,b)s(Y,b)$. By 7.12 this implies

$$\|v(X)g(Y) - v(Y)g(X)\| \subseteq U.$$

since $M'$ is binary. But the expression on the left is non-null by Axiom I, and since $g(U)$ is an elementary

chain of $N'$ it follows that

$$g(U)=\pm(v(X)g(Y)-v(Y)g(X)).$$

Hence, by the definitions of $f_U$, $f_V$, and $f_W$, we have

$$f_U=\pm(f_V(a)f_W-f_W(a)f_V).$$

If instead $U$ is the line $L'$ of $M'$ we have $\theta U=U\cup\{a\}=L$. Then the preceding equation follows at once from the definition of $f_U$. The required linear dependence is thus established for the case $a\epsilon L$.

Now suppose $a\notin L$. It may happen that $L$ is a line of $M'$ so that $\theta L=L$. Then $U$, $V$, and $W$ are flats of $M'$, and therefore points of $M'$ on $L$, by 4.14. There is a linear relation between $g(U)$, $g(V)$, and $g(W)$, as we may see by applying 2.31 to the regular chain-group $N'\times L$. A corresponding relation holds between $f_U$, $f_V$, and $f_W$.

In the remaining case $L$ is a flat, but not a line, of $M'$, by 4.45. Then $\theta L=L\cup\{a\}$, and $\theta L$ is a plane of $M$. It is a connected plane since some point $X$ of $M'$ on $L$ must satisfy $a\epsilon\theta X$. (See sec 4.4.) By 4.26 there are connected lines $L_U$, $L_V$ and $L_W$ on $\theta L$, distinct from $L$, which are on $U$, $V$, and $W$ respectively.

We can choose these lines to have no common point. For suppose $T$ is such a common point. There is a point $T'$, distinct from $T$ and $U$, on $L_U$. By 4.26 there is a connected line $L'$, distinct from $L_U$, on $\theta L$ and $T'$. Without loss of generality we may suppose $L'$ to meet $L$ in $V$, by 5.35. Replacing $L_V$ by $L'$ we obtain three lines of the kind required.

Write $X=\langle L_V\cap L_W\rangle$, $Y=\langle L_W\cap L_U\rangle$ and $Z=\langle L_U\cap L_V\rangle$. We note that $\langle\theta L-\{a\}\rangle=L$, by 7.11. Hence $a\epsilon X\cap Y\cap Z$. We may therefore apply the result already proved to $L_U$, $L_V$, and $L_W$, and obtain

$$f_U=\pm(f_Y(a)f_Z-f_Z(a)f_Y),$$

$$f_V=\pm(f_Z(a)f_X-f_X(a)f_Z),$$

$$f_W=\pm(f_X(a)f_Y-f_Y(a)f_X).$$

Thus $f_U$, $f_V$, and $f_W$ are linearly dependent in this case also. This completes the proof of 7.415.

The preceding result shows that the chains $f_U$ corresponding to the atoms of $M$ satisfy the conditions of 5.11. They therefore generate an integral chain-group $N$ such that $M=M(N)$. But the coefficients in each $f_U$ are restricted to the values 1, $-1$ and 0. Hence the $f_U$ are primitive chains of $N$. It follows that $N$ is a regular chain-group and $M$ is a regular matroid. We have thus completed the proof of 7.41.

### 7.5. Characterization of Regular Matroids

Combining the results 5.53, 7.21, 7.31, and 7.41 we have the theorem.

7.51 *A binary matroid $M$ is regular if and only if it has no minor of Type BI or BII.*

An equivalent condition is that $M$ shall include no Fano configuration or heptahedron.

We may also apply 5.26 and 5.23 to obtain the following rule.

7.52 *A binary matroid $M$ is non-regular if and only if some standard representative matrix $K$ of $M$ has a submatrix $J$ such that either $J$ or its transpose is of the following form, to within a permutation of columns.*

$$\begin{Bmatrix}1&0&1&1\\1&1&0&1\\0&1&1&1\end{Bmatrix}$$

In one respect the preceding theory is incomplete: we still need a convenient algorithm for determining whether a given binary matroid is regular.

## 8. The Matroid and the Atom

### 8.1. General Matroid

In this chapter we suppose given a matroid $M$ on a set $E$, and we fix some atom $Y$ of $M$. We discuss theorems concerned with the relationship of $Y$ to the rest of the matroid.

We refer to the elementary separators of $M\cdot(E-Y)$ as the *bridges* of $Y$ in $M$. Such a bridge is *trivial* if it contains no atom of $M\cdot(E-Y)$, and *monatomic* if it contains just one. A trivial bridge of $Y$ consists of a single cell of $E-Y$ contained in no atom of $M\cdot(E-Y)$, and every such cell of $E-Y$ defines a trivial bridge of $Y$. A nontrivial bridge of $Y$ in $M$ is thus a flat of $M\cdot(E-Y)$.

8.11 $d(M\cdot(E-Y))=dM-1$.

PROOF: Since $d(M\times Y)=dY=0$ this follows from 3.54

8.12 *Suppose $S\subseteq E-Y$ and let $d$ be an integer $\geqslant -1$. Then $S$ is a $d$-flat of $M\cdot(E-Y)$ if and only if $S\cup Y$ is a $(d+1)$-flat of $M$.*

PROOF: Suppose $S$ is a flat of $M\cdot(E-Y)$. Then each atom of $M\cdot(E-Y)$ on $S$ is the intersection with $E-Y$ of an atom of $M$ contained in $S\cup Y$. Hence $S\cup Y$ is a union of atoms of $M$, that is a flat of $M$.

Conversely suppose $S\cup Y$ is a flat of $M$. Each cell of $S$ belongs to an atom of $M$ on $S\cup Y$ and therefore to an atom of $M\cdot(E-Y)$ on $S$, by 1.11 and the definition of a reduction. Hence $S$ is a flat of $M\cdot(E-Y)$.

To complete the proof we observe that

$$d((M\cdot(E-Y))\times S)=d((M\times(S\cup Y))\cdot S)$$

$$=d(M\times(S\cup Y))-1,$$

by 3.333 and 8.11.

8.13 *Let $S$ be a connected flat of $M\cdot(E-Y)$. Then the only possible non-trivial separators of $M\times(S\cup Y)$ are $S$ and $Y$. Moreover $S$ and $Y$ are separators of $M\times(S\cup Y)$ if and only if*

$$(M\cdot(E-Y))\times S=M\times S.$$

PROOF: Let $Z$ be a separator of $M \times (S \cup Y)$. Then either $Y \subseteq Z$ or $Y \cap Z = \phi$. Moreover $Z \cap S$ is a separator of $(M \times (S \cup Y)) \cdot S$, by 3.43, and this matroid is $(M \cdot (E-Y)) \times S$, by 3.334. Hence $S \subseteq Z$ or $S \cap Z = \phi$, by hypothesis. Accordingly $Y$ and $S$ are the only possible nontrivial separators of $M \times (S \cup Y)$.

The necessary and sufficient condition for $S$ and $Y$ to be separators of $M \times (S \cup Y)$ is

$$(M \times (S \cup Y)) \cdot S = (M \times (S \cup Y)) \times S,$$

by 3.42. This is equivalent to the condition stated in the enunciation, by 3.331 and 3.333.

In what follows we shall be concerned only with binary matroids, and we do not pursue the general theory any further.

## 8.2. Binary Matroid

From now on we suppose $M$ binary. If $Z \epsilon M \cdot (E-Y)$ then $Y \cup Z$ is a line of $M$, by 8.12. If it is connected it has just two points other than $Z$, by 5.35. Their intersections with $Y$ are complementary non-null subsets $T$ and $U$ of $Y$, by 7.12. We call these subsets the *primary segments* of $Y$ determined by $Z$. The unordered pair $\{T, U\}$ is the *partition* of $Y$ determined by $Z$. If $Y \cup Z$ is not connected it has only two points $Y$ and $Z$. In this case it is convenient to say that the partition of $Y$ determined by $Z$ is $\{Y, \phi\}$, and that the corresponding primary segments are $Y$ and $\phi$.

Let $B$ be any bridge of $Y$ in $M$. If there is an atom of $M \cdot (E-Y)$ on $B$ we make the following definition. A *segment of $Y$ determined by $B$*, briefly a "$B$-segment" of $Y$ is a minimal non-null intersection of primary segments of $Y$ determined by atoms of $M \cdot (E-Y)$ on $B$. If $B$ is trivial we say that the only $B$-segment of $Y$ is $Y$ itself. Evidently we have

8.21 *The $B$-segments of $Y$ are disjoint non-null subsets of $Y$ whose union is $Y$.*

The class of $B$-segments of $Y$ is the *partition* of $Y$ determined by $B$. We denote it by $\pi(M, B, Y)$.

Two bridges $B_1$ and $B_2$ of $Y$ in $M$ will be said to *avoid* one another if there exists $S_1 \epsilon \pi(M, B_1, Y)$ and $S_2 \epsilon \pi(M, B_2, Y)$ such that $S_1 \cup S_2 = Y$, and to *overlap* one another in the contrary case.

The atom $Y$ is *bridge-separable* in $M$ if its bridges can be classified in two disjoint classes $P$ and $Q$ so that no two bridges belonging to the same class overlap. In an extreme case $Y$ may have no two bridges which overlap. We then say $Y$ is *totally* bridge-separable in $M$.

We call a binary matroid $M$ an *even* matroid if every atom is bridge-separable.

The following examples may clarify our terminology.

Consider a binary matroid of Type BI. This is the matroid of a binary chain-group $N$ which, with an appropriate enumeration of cells, has the following standard representative matrix.

$$K = \begin{Bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{Bmatrix}$$

Let us denote the cell corresponding to the $j^{\text{th}}$ column by $a_j$.

We take $Y$ to be the atom $\{a_3, a_5, a_6, a_7\}$ corresponding to the third row of $K$. Then $E-Y = \{a_1, a_2, a_4\}$. (See 5.22.)

It is clear that the only nonzero chains of $N \cdot (E-Y)$ are those with representative vectors (1, 0, 1), (0, 1, 1) and their sum (1, 1, 0). Moreover these three chains are elementary. Accordingly the matroid $M \cdot (E-Y) = M(N \cdot (E-Y))$ has just three atoms $\{a_1, a_2\}$, $\{a_2, a_4\}$ and $\{a_1, a_4\}$. (See 3.31.)

We observe that $M \cdot (E-Y)$ is connected. Hence $\{a_1, a_2, a_4\}$ is the only bridge of $Y$ in $M$. Accordingly $Y$ must be classed as bridge-separable.

It can be verified that all seven atoms of $M$ are bridge-separable. Indeed they are all equivalent under the symmetry of the Fano configuration. Hence the matroids of Type BI are even.

In our example the sets $\{a_1, a_4\} \cup Y$ and $\{a_2, a_4\} \cup Y$ are lines of $M$, by 8.12. They are on points corresponding to the first and second rows of $K$ respectively. We deduce that they determine the partitions $\{\{a_3, a_5\}, \{a_6, a_7\}\}$ and $\{\{a_3, a_6\}, \{a_5, a_7\}\}$ of $Y$ respectively.

Considering the intersections of these four primary segments we see that

$$\pi(M, B, Y) = \{\{a_3\}, \{a_5\}, \{a_6\}, \{a_7\}\}.$$

Next we discuss binary matroids of Type BII. Such a matroid $M$ corresponds to a binary chain-group $N$ with the following representative matrix $K$, by 5.23.

$$K = \begin{Bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{Bmatrix}$$

Again we enumerate the cells as $a_1, a_2, \ldots, a_7$, in the order of the columns of $K$.

We take $Y$ to be the atom $\{a_4, a_5, a_6, a_7\}$ corresponding to the last row of $K$. Now each of the sets $\{a_1\}$, $\{a_2\}$ and $\{a_3\}$ are domains of chains of $N \cdot (E-Y)$. Hence these three sets are the only atoms of $M \cdot (E-Y)$, by Axiom I. Accordingly they are also the elementary separators of $M \cdot (E-Y)$, that is they are the three bridges of $Y$ in $M$.

The line $Y \cup \{a_1\}$ of $M$ is on a point corresponding to the first row of $M$. We deduce that the atom $\{a_1\}$ of $M \cdot (E-Y)$ determines the partition $\{\{a_4, a_7\}, \{a_5, a_6\}\}$ of $Y$. Since $\{a_1\}$ is a monatomic bridge of

$Y$ in $M$ we thus have

$$\pi(M, \{a_1\}, Y) = \{\{a_1, a_7\}, \{a_5, a_6\}\}.$$

Similarly

$$\pi(M, \{a_2\}, Y) = \{\{a_4, a_5\}, \{a_6, a_7\}\},$$

$$\pi(M, \{a_3\}, Y) = \{\{a_4, a_6\}, \{a_5, a_7\}\}.$$

Inspection of these formulae shows that the three bridges of $Y$ in $M$ are mutually overlapping. Hence $Y$ is not bridge-separable and the binary matroid $M$ is *odd*, that is not even.

8.22 *Every binary matroid of Type BII is odd.*

## 8.3. Kuratowski Graphs

We now show the application of our terminology to certain matroids associated with graphs.

Suppose first that $G$ is the complete 5-graph shown in figure 5B. We use the notation for it given at the end of chapter 5.

Let $M$ be the polygon-matroid of $G$, so that atoms of $M$ correspond to polygons of $G$. In particular there is an atom

$$Y = \{L_{12}, L_{23}, L_{34}, L_{45}, L_{15}\}$$

of $M$ corresponding to the outer polygon in figure 5B.

Now $M \cdot (E-Y)$ is the polygon matroid of $G \times (E-Y)$, by 3.372. In this graph the polygon $G \cdot Y$ has been replaced by a single vertex $v$, and each cell of $E-Y$ has become a loop whose ends coincide in $v$. Accordingly each cell of $E-Y$ constitutes by itself an atom of $M \cdot (E-Y)$. It also constitutes an elementary separator of $M \cdot (E-Y)$, by Axiom I. We deduce that there are just five bridges of $Y$ in $M$, namely the sets

$$\{L_{13}\}, \{L_{24}\}, \{L_{35}\}, \{L_{14}\}, \{L_{25}\}.$$

Considering the triangle defined by $L_{12}$, $L_{23}$, and $L_{13}$ we see that $\{L_{13}\}$ determines the partition $\{\{L_{12}, L_{23}\}, \{L_{34}, L_{45}, L_{15}\}\}$ of $Y$. Hence, using the symmetry of $G$, we have

$$\pi(M, \{L_{13}\}, Y) = \{\{L_{12}, L_{23}\}, \{L_{34}, L_{45}, L_{15}\}\},$$

$$\pi(M, \{L_{24}\}, Y) = \{\{L_{23}, L_{34}\}, \{L_{45}, L_{15}, L_{12}\}\},$$

and so on.

Inspection of these formulae shows that $\{L_{13}\}$ overlaps $\{L_{24}\}$, $\{L_{24}\}$ overlaps $\{L_{35}\}$, $\{L_{35}\}$ overlaps $\{L_{14}\}$, $\{L_{14}\}$ overlaps $\{L_{25}\}$ and $\{L_{25}\}$ overlaps $\{L_{13}\}$. It is thus impossible to arrange these five bridges in two classes so that no two members of the same class overlap. Thus $Y$ is not bridge-separable.

Next let $G$ be a Thomsen graph as shown in the second diagram of figure 5B. A similar analysis applies.

We write $M = P(G)$ and take $Y = \{L_{11}, L_{21}, L_{22}, L_{32}, L_{33}, L_{13}\}$, corresponding to the outer polygon of the diagram. Then we find that $\{L_{12}\}$, $\{L_{23}\}$, and $\{L_{31}\}$ are the three atoms of $M \cdot (E-Y)$ and the three bridges of $Y$ in $M$. Considering the quadrilaterals such as $a_1\, b_1\, a_2\, b_2$ we arrive at the following partition formulae.

$$\pi(M, \{L_{12}\}, Y) = \{\{L_{11}, L_{21}, L_{22}\}, \{L_{32}, L_{33}, L_{13}\}\},$$

$$\pi(M, \{L_{23}\}, Y) = \{\{L_{22}, L_{32}, L_{33}\}, \{L_{13}, L_{11}, L_{21}\}\},$$

$$\pi(M, \{L_{31}\}, Y) = \{\{L_{33}, L_{13}, L_{11}\}, \{L_{21}, L_{22}, L_{32}\}\}.$$

We see from these formulae that the three bridges are mutually overlapping. Hence $Y$ is not bridge-separable.

We say that a matroid is of Type KI or KII if it can be interpreted as the polygon-matroid of a complete 5-graph or a Thomsen graph respectively. We sum up the results of the present section in the following theorem

8.31 *Every matroid of Type KI or KII is odd.*

## 8.4. Skewness

In this section $M$ denotes an arbitrary binary matroid and $Y$ denotes an atom of $M$.

We say that an atom $Z$ of $M \cdot (E-Y)$ *cuts* a subset $S$ of $Y$ if it determines a partition $\{T, U\}$ of $Y$ such that $S \cap T$ and $S \cap U$ are both non-null.

8.41 *The points of $M \cdot (E-Y)$ which do not cut a given subset $S$ of $Y$ constitute a linear subclass of $M \cdot (E-Y)$.*

PROOF: Let $Z_1$, $Z_2$, and $Z_3$ be the three points on some connected line $L$ of $M \cdot (E-Y)$. Suppose $Z_1$ and $Z_2$ do not cut $S$. Then each point on the lines $Y \cup Z_1$ and $Y \cup Z_2$ of $M$ either contains $S$ or does not meet $S$. That is it does not "cut" $S$ in the sense of section 4.3.

Let $X_3$ be any point of $M$, other than $Y$, on the line $Y \cup Z_3$. There is a line $L'$ of $M$ on $X_3$ and the plane $Y \cup L$ which is not on $Y$, by 4.15. This line meets $Y \cup Z_1$ and $Y \cup Z_2$ in two distinct points of $M$, by 4.171. Hence $X_3$ either contains $S$ or does not meet $S$, by 4.31. It follows that the point $Z_3$ of $M \cdot (E-Y)$ does not cut $S$.

We say that two points $Z$ and $Z'$ of $M \cdot (E-Y)$ are *skew* with respect to $Y$ if they determine partitions $\{T, U\}$ and $\{T', U'\}$ of $Y$ such that the intersections $T \cap T'$, $T \cap U'$, $U \cap T'$ and $U \cap U'$ are all non-null.

We extend this definition by saying that two bridges $B$ and $B'$ of $Y$ in $M$ are *skew* if there are points $Z$ and $Z'$ of $M \cdot (E-Y)$, skew with respect to $Y$, which are on $B$ and $B'$ respectively. We also say that a point $Z$ of $M \cdot (E-Y)$ is *skew* to a bridge $B$ of $Y$ in $M$ if there is a point $Z'$ of $M \cdot (E-Y)$ on $B$ which is skew to $Z$.

8.42 *Let $B_1$, $B_2$ and $B_3$ be distinct bridges of $Y$ in $M$ such that $B_2$ is skew both to $B_1$ and to $B_3$. Then either there is a point $Z_2$ of $M \cdot (E-Y)$ on $B_2$ which is skew to both $B_1$ and $B_3$ or there are points $Z_1$ and $Z_3$ of $M \cdot (E-Y)$ on $B_1$ and $B_3$ respectively such that $Z_1$, $B_2$, and $Z_3$ are mutually skew.*

PROOF: There are points $Z_1$, $Z_2'$, $Z_2''$, and $Z_3$ of $M \cdot (E-Y)$, on $B_1$, $B_2$, $B_2$, and $B_3$ respectively, such that $Z_1$ is skew to $Z_2'$ and $Z_2''$ is skew to $Z_3$. If $Z_1$ and

$Z_3$ are skew the second alternative of the theorem holds.

In the remaining case $Z_1$ and $Z_3$ determine partitions $\{S_1, T_1\}$ and $\{S_3, T_3\}$ of $Y$ such that $T_1 \cap T_3 = \phi$, that is $T_1 \subseteq S_3$ and $T_3 \subseteq S_1$.

Now $Z_2'$ cuts $T_1$, $Z_2''$ cuts $T_3$, and $B_2$ is a connected flat of $M \cdot (E-Y)$. Hence, by 4.37 and 8.41 there is a point $Z_2$ of $M \cdot (E-Y)$ on $B_2$ cutting both $T_1$ and $T_3$. But then $Z_2$ is skew to both $Z_1$ and $Z_3$, that is to both $B_1$ and $B_3$.

Let $B$ be any bridge of $Y$ in $M$. We call it an *n-bridge* of $Y$, where $n$ is an integer, if

$$|\pi(M, B, Y)| = n.$$

Two $n$-bridges $B_1$ and $B_2$ of $Y$ in $M$ are *equipartite* if

$$\pi(M, B_1, Y) = \pi(M, B_2, Y)$$

8.43 *Let $B$ be a 2-bridge or 3-bridge of $Y$ in $M$, and suppose $W \in \pi(M, B, Y)$. Then there is a point $Z$ of $M \cdot (E-Y)$ on $B$ which determines the partition $\{W, Y-W\}$ of $Y$.*

PROOF: If $B$ is a 2-bridge this result follows at once from the definitions. We may therefore write $\pi(M, B, Y) = \{U, V, W\}$. There is a point of $M \cdot (E-Y)$ on $B$ which cuts $U \cup W$ and another which cuts $V \cup W$. Hence, by 4.37 and 8.41 there is such a point $Z$ which cuts both. This can only determine the partition $\{W, U \cup V\}$ of $Y$.

The next theorem shows the relation between skewness and overlapping for bridges.

8.44 *Let $B_1$ and $B_2$ be overlapping bridges of $Y$ in $M$. Then either $B_1$ and $B_2$ are skew or they are equipartite 3-bridges.*

PROOF: Assume $B_1$ and $B_2$ are not skew.

Let $Z_1$ be a point of $M \cdot (E-Y)$ on $B_1$ determining a partition $\{S, T\}$ of $Y$. Suppose $S$ and $T$ each meet more than one member of $\pi(M, B_2, Y)$. Then $S$ and $T$ are each cut by some point of $M \cdot (E-Y)$ on $B_2$, and therefore there is one such point which cuts both of them, by 4.37 and 8.41. But then $B_1$ and $B_2$ are skew, contrary to assumption. We deduce that one of $S$ and $T$ must be contained in a member of $\pi(M, B_2, Y)$.

Suppose $B_1$ is a 2-bridge or 3-bridge. Then if $W \in \pi(M, B_1, Y)$ we can apply the preceding result with $W = S$ and $Y - W = T$, by 8.43. We find that there exists $U \in \pi(M, B_2, Y)$ such that either $W \subseteq U$ or $Y - W \subseteq U$. The second alternative must be ruled out since it implies $W \cup U = Y$, contrary to the hypothesis that $B_1$ and $B_2$ overlap. We deduce further that $B_1$ is not a 2-bridge since then $Y - W \in \pi(M, B_1, Y)$ and both alternatives are ruled out.

Suppose $B_1$ is a 3-bridge. Write $\pi(M, B_1, Y) = \{U, V, W\}$. Then by the above reasoning there are members $U'$, $V'$, and $W'$ of $\pi(M, B_2, Y)$ such that $U \subseteq U'$, $V \subseteq V'$, and $W \subseteq W'$. These three $B_2$-segments are distinct. If for example we had $U' = V'$ it would follow that $U' \cup W = Y$, contrary to hypothesis. It follows that $B_1$ and $B_2$ are equipartite 3-bridges.

We have still to consider the case in which $B_1$ is an $n$-bridge, with $n \geq 4$.

Let $W_1$, $W_2$, $W_3$, and $W_4$ be distinct members of $\pi(M, B_1, Y)$. Using 4.37 and 8.41 we find that there is a point $Z$ of $M \cdot (E-Y)$ on $B_1$ which cuts both $W_1 \cup W_2$ and $W_3 \cup W_4$. It determines a partition $\{S, T\}$ of $Y$ such that $W_1$ and $W_2$ are in different sets $S$ and $T$, as are also $W_3$ and $W_4$.

By the first part of the proof we can find $U' \in \pi(M, B_2, Y)$ such that $U'$ contains two of the given four $B_1$-segments. Let $W'$ be a member of $\pi(M, B_2, Y)$ containing the greatest possible number $k$ of $B_1$-segments. $(k \geq 2)$.

Suppose there are at least two $B_1$-segments not in $W'$. Then by the result just proved for four $B_1$-segments $W_1$, $W_2$, $W_3$, $W_4$ it follows that some point of $M \cdot (E-Y)$ on $B_1$ determines a partition $\{S, T\}$ of $Y$ such that $S$ and $T$ each meet both $W'$ and $Y - W'$. Since some member of $\pi(M, B_2, Y)$ must contain $S$ or $T$ we have a contradiction. We are forced to conclude that there is at most one $B_1$-segment which is not contained in $W'$. But then $B_1$ and $B_2$ avoid one another, contrary to hypothesis.

This completes the proof of the theorem.

## 8.5. Segments Determined by Bridges

This section deals with some auxiliary results to be used in later proofs. They enable us to go from $Y$ to another atom of the binary matroid $M$, or to replace $M$ by one of its minors.

8.51 *Let $S$ be a subset of $E$, and let $X$ be an atom of $M$ such that $X \cap S$ is non-null. Then $X \cap S$ is a union of disjoint atoms of $M \cdot S$.*

PROOF: Let $N$ be the binary chain-group such that $M = M(N)$. Then $M \cdot S = M(N \cdot S)$, by 3.31. But $X \cap S$ is the domain of a chain of $N \cdot S$. The theorem now follows from 5.32.

8.52 *Let $B$ be a bridge of $Y$ in $M$. Then the $B$-segments of $Y$ in $M$ are the minimal non-null intersections of atoms of $(M \times (B \cup Y)) \cdot Y$.*

PROOF. Let the class of minimal non-null intersections of atoms of $(M \times (B \cup Y)) \cdot Y$ be denoted by $J$. The members of $J$ are disjoint and their union is $Y$, by 1.11.

Let $a$ be any cell of $Y$. It is contained in a unique $B$-segment $U$ of $Y$ and a unique member $W$ of $J$.

Consider an atom $Z$ of $M \cdot (E-Y)$ on $B$ determining a partition $\{S, T\}$ of $Y$, where $a \in S$. Since $S \cup Z$ is an atom of $M \times (B \cup Y)$, by 8.12, unless $S = Y$, it follows that $W \subseteq S$. Since this result must hold for all choices of $Z$ we have $W \subseteq U$.

Suppose $W$ is not the whole of $U$. Then there is an atom $X$ of $M \times (B \cup Y)$ which cuts $U$, in the sense of section 4.3. Choose such an $X$ so that $|B \cap X|$ has the least possible value.

Assume $B \cap X$ is not an atom of $(M \times (B \cup Y)) \cdot B$. Then there is an atom $Z$ of this matroid such that $Z \subseteq B \cup X$, and there is an atom $X'$ of $M \times (B \cup Y)$ such that $X' \cap B = Z$. By the choice of $X$ the set $X'$ does not cut $U$.

Consider the mod 2 sum $X + X'$. Clearly it cuts $U$. But it is a union of disjoint atoms of $M \times (B \cup Y)$, by

5.34, and one of these must cut $U$. This is contrary to the choice of $X$.

We deduce that $B \cap X$ is in fact an atom of $(M \times (B \cup Y)) \cdot B$, that is $(M \cdot (E-Y)) \times B$, by 3.334. Hence $B \cap X$ is an atom of $M \cdot (E-Y)$ on $B$. But then $X$ does not cut $U$, by the definition of a $B$-segment.

From this contradiction it follows that in fact $W = U$. Since this result holds for all choices of $a$ the theorem follows.

8.53 *Let $B$ be a bridge of $Y$ in $M$, and $S$ a subset of $E$ such that $B \cup Y \subseteq S$. Then $B$ is a bridge of $Y$ in $M \times S$, and*

$$\pi(M \times S, B, Y) = \pi(M, B, Y).$$

PROOF: We have

$$(M \times S) \cdot (S-Y) = (M \cdot (E-Y)) \times (S-Y),$$

by 3.334. Hence $B$ is a separator of $(M \times S) \cdot (S-Y)$, by 3.43. Moreover

$$((M \cdot (E-Y)) \times (S-Y)) \times B = (M \cdot (E-Y)) \times B,$$

by 3.331, and this matroid is connected. Hence $B$ is a bridge of $Y$ in $M \times S$. Since

$$(M \times S) \times (B \cup Y) = M \times (B \cup Y),$$

by 3.331, the theorem now follows from 8.52.

8.531 *Let $B$ be a bridge of $Y$ in $M$. Let $S$ be a subset of $E$ such that $B \cup Y \subseteq S$ and no $Z \in M \cdot (E-Y)$ is a subset of $E-S$. Then $Y \in M \cdot S$, $B$ is a bridge of $Y$ in $M \cdot S$, and*

$$\pi(M \cdot S, B, Y) = \pi(M, B, Y).$$

PROOF: We have

$$(M \cdot S) \cdot (S-Y) = (M \cdot (E-Y)) \cdot (S-Y),$$

by 3.332. Hence $B$ is a separator of $(M \cdot S) \cdot (S-Y)$, by 3.43. Moreover

$$((M \cdot S) \cdot (S-Y)) \cdot B = (M \cdot (E-Y)) \cdot B,$$

by 3.332, and this matroid is connected. Hence $B$ is a bridge of $Y$ in $M \cdot S$, provided that $Y$ is an atom of $M \cdot S$.

Suppose some atom $X$ of $M \times ((E-S) \cup B \cup Y)$ meets $E-S$. Then some $Z \in M \cdot (E-Y)$ meets $E-S$ and satisfies

$$Z \subseteq X \cap (E-Y) \subseteq (E-S) \cup B,$$

by 1.11. But then $Z \subseteq E-S$, since $B$ is a separator of $M \cdot (E-Y)$, and this is contrary to the definition of $S$. We deduce that

$$(M \cdot S) \times (B \cup Y)$$

$$= (M \times ((E-S) \cup B \cup Y)) \cdot (B \cup Y),$$

by 3.333,

$$= (M \times (B \cup Y)) \cdot (B \cup Y)$$

$$= M \times (B \cup Y).$$

We note that this implies $Y \in M \cdot S$. An application of 8.52 completes the proof.

8.54 *Let $B$ be any bridge of $Y$ in $M$. Let $S$ be a subset of $E$ such that $E-Y \subset S$. Then $S \cap Y \in M \cdot S$, $B$ is a bridge of $S \cap Y$ in $M \cdot S$, and*

$$\pi(M \cdot S, B, S \cap Y)$$

*is the class of all non-null intersections with $S$ of members of $\pi(M, B, Y)$.*

PROOF: There exists $Y' \in M \cdot S$ such that $Y' \subseteq S \cap Y$. Moreover there exists $Y'' \in M$ such that $Y'' \cap S = Y'$. But then $Y'' \subseteq Y$. Hence $Y'' = Y$ and $S \cap Y = Y' \in M \cdot S$, by Axiom I.

We have

$$(M \cdot S) \cdot (S-(S \cap Y)) = M \cdot (S-(S \cap Y)) = M \cdot (E-Y),$$

by 3.332. Hence $B$ is a bridge of $S \cap Y$ in $M \cdot S$. Moreover

$$((M \cdot S) \times (B \cup (S \cap Y))) \cdot (S \cap Y)$$

$$= (M \times (E-(S-(B \cup Y)))) \cdot S \cap Y)$$

$$= ((M \times (B \cup Y)) \cdot Y) \cdot (S \cap Y),$$

by 3.332 and 3.333. The theorem now follows from 8.51 and 8.52.

8.55 *Let $B_1$ and $B_2$ be distinct bridges of $Y$ in $M$. Let $Z$ be a point of $M \cdot (E-Y)$ on $B_1$ determining a partition $\{S, T\}$ of $Y$ where $S$ and $T$ are non-null. Let $Y'$ be the point $S \cup Z$ of $M'$ Then the following propositions hold.*

(i) *$T$ is an atom of $M \cdot (E-Y')$.*

(ii) *There is a bridge $B'$ of $Y'$ in $M$ such that $B_2 \subseteq B'$.*

(iii) *If $T$ is contained in some $W \in \pi(M, B_2, Y)$, then $B_2 = B'$. Moreover $\pi(M, B', Y')$ differs from $\pi(M, B_2, Y)$ only in the replacement of $W$ by $(W-T) \cup Z$.*

(iv) *If $T$ is contained in no member of $\pi(M, B_2, Y)$, then $T \subseteq B'$. Moreover each member of $\pi(M, B', Y')$ is then contained either in $Z$ or in a member of $\pi(M, B_2, Y)$.*

PROOF: Since $Z \cup T$ is an atom of $M$ there is an atom $T'$ of $M \cdot (E-Y')$ such that $T' \subseteq T$. There is an atom $X$ of $M$ such that $X \cap (E-Y') = T'$. But then $X$ is a point of $M$ on the line $Y' \cup T = Y \cup Z$. Since the only points on this line are $Y$, $S \cup Z$ and $T \cup Z$ we have $X = T \cup Z$ and $T' = T$. This establishes (i).

We have

$$(M \cdot (E-Y')) \cdot B_2 = (M \cdot (E-Y)) \cdot B_2,$$

by 3.332. As the expression on the right represents a connected matroid we deduce that $B_2$ is contained

in some elementary separator $B'$ of $M\cdot(E-Y')$, by 3.46. Then $B'$ is a bridge of $Y'$ in $M$ such that $B_2\subseteq B'$. This establishes (ii).

To prove (iii) assume that $T$ is contained in some $W\epsilon\pi(M, B_2, Y)$. Let $Z_1$ be any atom of $M\cdot(E-Y')$ on $B'$ which meets $B_2$.

There is an atom $X$ of $M\times(B'\cup Y)$ such that $X\cap B'=Z_1$. Thus $X$ is an atom of $M$ meeting $B_2$. It follows that there is an atom $Z_2$ of $M\cdot(E-Y)$ such that $Z_2\cap B_2$ is non-null and $Z_2\subseteq X\cap(E-Y)$, by 1.11. Clearly $Z_2$ is on $B_2$. We thus have

$$Z_2\subseteq Z_2\cap B'\subseteq X\cap B'=Z_1.$$

$Z_2$ determines a partition $\{S_2, T_2\}$ of $Y$, and by our assumption we may suppose $T\subseteq T_2$. The set $S_2\cup Z_2$ is an atom of $M$ which meets $B_2$ and therefore $B'$. Hence there is an atom $Z_3$ of $M\cdot(E-Y')$ on $B'$ such that $Z_3\subseteq(Z_2\cup S_2)\cap(E-Y')$. But $S_2$ does not meet $E-Y'$ since $T\subseteq T_2$. Hence $Z_3\subseteq Z_2\subseteq Z_1$.

Applying Axiom I to $M\cdot(E-Y')$ we find that $Z_1=Z_2=Z_3$. We deduce that $B_2$ is a separator of $(M\cdot(E-Y'))\times B'$. Hence $B_2=B'$, by 3.44. We note further that

$$\begin{aligned}&(M\cdot(E-Y'))\times B_2\\&=(M\cdot(E-Y'))\cdot B_2\\&=(M\cdot(E-Y))\cdot B_2\\&=(M\cdot(E-Y))\times B_2,\end{aligned}$$

by 3.332 and 3.42. Hence $M\cdot(E-Y)$ and $M\cdot(E-Y')$ have the same atoms on $B_2$.

If one such atom $Z_0$ determines a partition $\{S_0, T_0\}$ of $Y$, with $T\subseteq T_0$, then it determines the partition $\{S_0, (T_0-T)\cup Z\}$ of $Y'$. For the point $S_0\cup Z_0$ is on both $Y\cup Z_0$ and $Y'\cup Z_0$. These results establish (iii).

Now let us assume that $T$ is contained in no member of $\pi(M, B_2, Y)$, and discuss proposition (iv).

There is an atom of $M\times(B_2\cup Y)$ which cuts $T$ in the sense of section 4.3. Hence there is an atom $Z_1$ of $M\cdot(E-Y')$ which cuts $T$ in this sense and satisfies $Z_1\subseteq(B_2\cup Y)\cap(E-Y')\subseteq B'\cup T$. Since $T\epsilon M\cdot(E-Y')$, by (i), it follows that $Z_1$ meets both $B'$ and $T$. Hence $T\subseteq B'$, by 3.48.

Now $Z$ is the intersection of $Y'$ with an atom $Z\cup T$ of $M\times(B'\cup Y')$. Hence it is a union of members of $\pi(M, B', Y')$ by 8.51 and 8.52.

Again, let $Z_0$ be any atom of $M\cdot(E-Y)$ on $B_2$, determining a partition $\{S_0, T_0\}$ of $Y$. Then $S\cap S_0$ is the intersection with $Y'$ of the atom $S_0\cup Z_0$ of $M\times(B'\cup Y')$, since $T\subseteq B'$. Hence $S\cap S_0$ and similarly $S\cap T_0$ are unions of members of $\pi(M, B', Y)$, by 8.51 and 8.52. Since this is true for each choice of $Z_0$ proposition (iv) follows.

## 8.6. Regular Matroid

We conclude this chapter with some theorems about binary matroids with no minors of one of the types BI and BII. These are all true for regular matroids, by 7.51.

8.61 *Suppose $M$ has no minor of Type BI. Let $Z_1$ and $Z_2$ be points on a connected line $L$ of $M\cdot(E-Y)$. Then $Z_1$ and $Z_2$ are not skew with respect to $Y$.*

PROOF: $L\cup Y$ is a plane $P$ of $M$, by 8.12. It includes three distinct lines $Y\cup Z_1$, $Y\cup Z_2$, and $Y\cup Z_3$ of $M$, where $Z_3$ is the third point of $M\cdot(E-Y)$ on $L$, by 8.12.

Assume $Z_1$ and $Z_2$ skew. They determine partitions $\{S_1, T_1\}$ and $\{S_2, T_2\}$ of $Y$ respectively. There are cells $a\epsilon S_1\cap S_2$, $b\epsilon S_1\cap T_2$, $c\epsilon T_1\cap S_2$, and $d\epsilon T_1\cap T_2$. The flats $\langle P-\{a\}\rangle$, $\langle P-\{b\}\rangle$, $\langle P-\{c\}\rangle$, and $\langle P-\{d\}\rangle$ are lines on $P$ which are not on $Y$, by 4.13. It is easily seen that they are distinct. For example $\langle P-\{a\}\rangle$ is the only one which is on the two points $Z_1\cup T_1$ and $Z_2\cup T_2$ of $M$.

We see that there are seven distinct lines on $P$. Thus $M$ includes a Fano configuration, which is contrary to hypothesis, by 7.21.

8.62 *Suppose $M$ has no minor of Type BI. Let $B$ be any non-trivial bridge of $Y$ in $M$, and let $W$ be any member of $\pi(M, B, Y)$. Then there exists $Z\epsilon M\cdot(E-Y)$ on $B$ determining the partition $\{W, Y-W\}$ of $Y$.* (Compare 8.43).

PROOF: Let $a$ be any cell of $Y$. Let $U$ be a subset of $Y$ such that $a\epsilon U$, some $Z\epsilon M\cdot(E-Y)$ on $B$ determines the partition $\{U, Y-U\}$ of $Y$, and $|U|$ has the least value consistent with these conditions.

Assume there is a $W\epsilon\pi(M, B, Y)$ such that $W\cap U$ is a non-null proper subset of $U$. Then there exists $X\epsilon M\times(B\cup Y)$ such that $X\cap U$ is a non-null proper subset of $U$. Clearly $X\cup Y$ is a connected flat of $M\times(B\cup Y)$. By 4.31 and 4.34 we can find a point $X'$ of $M\times(B\cup Y)$ on $X\cup Y$ which cuts $U$ in the sense of section 4.3 and has the property that $X'\cup Y$ is a connected line of $M\times(B\cup Y)$. Then $X'-Y$ is a point of $M\cdot(E-Y)$ on $B$, by 8.12. It determines the partition $\{X'\cap Y, Y-X'\}$ of $Y$, and so cuts $U$ in the sense of section 8.4.

By 4.34 and 8.41 we can find $Z'\epsilon M\cdot(E-Y)$ on $B$ which cuts $U$ and is such that $Z\cup Z'$ is a connected line of $M\cdot(E-Y)$. By 8.61 $Z'$ determines a partition $\{S, Y-S\}$ of $Y$ such that $S\subset U$. By the definition of $U$ we have $a\epsilon U-S$.

Choose $b\epsilon S$. By 8.12 the set $Y\cup Z\cup Z'$ is a plane $P$ of $M$. The line $\langle P-\{b\}\rangle$ of $M$ is on the points $Z\cup(Y-U)$ and $Z'\cup(Y-S)$ of $M$. Since these are distinct and not disjoint their mod 2 sum, $Q$ say, is also a point of $M$ on $\langle P-\{b\}\rangle$, by 5.34 and 5.35. But the mod 2 sum of $Z$ and $Z'$ is the third point of $M\cdot(E-Y)$ on the connected line $Z\cup Z'$. It determines the partition

$$\{Q\cap Y, Y-Q\}=\{U-S, Y-(U-S)\}$$

of $Y$. But this contradicts the definition of $U$.

We deduce from this contradiction that $U$ is itself a member of $\pi(M, B, Y)$. Since $a$ may be any cell of $Y$ the theorem follows.

8.63 *Suppose $M$ has no minor of Type BI, and that $dM\geq 2$. Then there is an atom $Y$ of $M$ which has two distinct bridges in $M$.*

PROOF: Assume the theorem false. Choose $X \epsilon M$. Then $E-X$ is a bridge of $X$ in $M$. It is nontrivial since $dM \geq 2$.

Choose $W \epsilon \pi(M, E-X, X)$. There exists $Z \epsilon M \cdot (E-X)$ determining the partition $(W, X-W)$ of $X$, by 8.62. Then $X \cup Z$ is a line of $M$, possibly disconnected, and $Z \cup (X-W), = Y$ say, is a point of $M$. Since $dM \geq 2$ the set $E-(X \cup Z)$ is non-null.

$W$ is a point of $M \cdot (E-Y)$, by 8.55. Suppose $Z' \epsilon M \cdot (E-Y)$ and $Z' \cap W$ is non-null. Then there exists $V \epsilon M$ such that $Z' = V \cap (E-Y)$. Then $V \cap W = Z' \cap W$. Hence $W \subseteq Z'$ since $W \epsilon \pi(M, E-X, X)$, and therefore $W = Z'$ by Axiom I.

It follows that $W$ is a separator of $M \cdot (E-Y)$. But $(E-Y)-W = E-(Y \cup Z) \neq \phi$. Hence there are at least two bridges of $X$ in $M$, one being $W$.

8.64 *Suppose $M$ has no minor of Type BII. Let $Z_1$, $Z_2$ and $Z_3$ be points of $M \cdot (E-Y)$ on distinct bridges $B_1$, $B_2$ and $B_3$ respectively of $Y$ in $M$. Let $Z_i$ determine the partition $\{S_i, T_i\}$ of $Y$. $(i = 1, 2, 3)$. Then if $S_3$ meets both $S_1 \cap S_2$ and $T_1 \cap T_2$ it contains one of the sets $S_1 \cap T_2$ and $S_2 \cap T_1$.*

PROOF: Suppose the theorem false. Then we can choose $K = \{a, b, c, d, e, f, g\} \subseteq E$ such that $a \epsilon Z_1$, $b \epsilon Z_2$, $c \epsilon Z_3$, $d \epsilon S_1 \cap S_2 \cap S_3$, $e \epsilon T_1 \cap T_2 \cap S_3$, $f \epsilon (S_1 \cap T_2) - S_3$, and $g \epsilon (S_2 \cap T_1) - S_3$.

In the matroid $M \cdot (E-Y)$ the sets $Z_1 \cup Z_2$, $Z_2 \cup Z_3$, and $Z_3 \cup Z_1$ are disconnected lines, by 4.24. Moreover $Z_1 \cup Z_2 \cup Z_3$ is a plane since there are only three disjoint points on it. ($\{a, b, c\}$ is a dendroid of $M \times (Z_1 \cup Z_2 \cup Z_3)$). Hence $Y \cup (Z_1 \cup Z_2 \cup Z_3)$ is a 3-flat $J$ of $M$, by 8.12.

There are points of $M$ on $J$ whose intersections with $K$ are $\{a, d, f\}$, $\{a, g, e\}$, $\{b, d, g\}$, $\{b, e, f\}$, $\{c, f, g\}$, $\{c, d, e\}$, and $\{a, b, c\}$. For example $(Z_1 \cup S_1) \cap K = \{a, d, f\}$, and similar verifications may be made for the next five intersections. As for $\{a, b, c\}$ we observe that

$$Z_1 \cup T_1 \subset Z_1 \cup T_1 \cup Z_2 \cup T_2 \subset Z_1 \cup T_1 \cup Z_2 \cup T_2 \cup Z_3 \cup T_3 \subset J.$$

Hence $Z_1 \cup T_1 \cup Z_2 \cup T_2$ is a line of $M$, by 4.14. Its points are $Z_1 \cup T_1$, $Z_2 \cup T_2$ and their mod 2 sum, $Q$ say. We have $Q \cap K = \{a, b, f, g\}$. Moreover $Z_1 \cup T_1 \cup Z_2 \cup T_2 \cup Z_3 \cup T_3$ is a plane $P$ of $M$ on $E$, by 4.14. The line $\langle P-\{e\} \rangle$ of $M$ is on the points $Z_3 \cup T_3$ and $Q$. As these two distinct points have a common cell $f$ their mod 2 sum, $R$ say, is the third point of $M$ on $\langle P-\{e\} \rangle$. But $R \cap K = \{a, b, c\}$.

Considering the seven intersections with $K$ set out above we see that each cell of $K$ occurs in just three of them and that no two have two cells in common. Hence given any three cells $x$, $y$, and $z$ of $K$ we can find a point of $M$ on $J$ which includes $x$ but not $y$ or $z$. It follows that no three of the planes $\langle J-\{x\} \rangle$, $x \epsilon K$, have a common line; we can find a point on any two of them which is not on the third.

Hence these seven planes on $J$ are distinct and define a heptahedron. This is contrary to hypothesis, by 7.31.

# 9. Graphic Matroids

## 9.1. Bonds

We consider the group $N$ of coboundaries of a graph $G$ over a commutative ring $R$, with a unit element and no divisors of zero. As explained in section 2.5 $M(N)$ is the bond-matroid $B(G)$ of $G$, and is independent of $R$.

9.11 *Let $f$ be any 0-chain on $G$ over $R$. Let $H$ be any component of the spanning subgraph $G: (E(G)-\|\delta f\|$ of $G$. (See sec. 3.1.) Then the vertices of $H$ have equal coefficients in $f$.*

PROOF: Suppose the contrary. Then some edge $A$ of $H$ has ends with different coefficients in $f$. But then $A \epsilon \|\delta f\|$, by the definition of a coboundary, which is contrary to the definition of $H$.

9.12 *Let $g$ be an elementary chain of $N$. Then there are distinct components $H$ and $K$ of $G: (E(G)-\|g\|)$ such that each edge of $\|g\|$ has one end in $V(H)$ and one in $V(K)$.*

PROOF: Write $g = \delta f$. Let $A$ be any edge of $\|g\|$. Its ends $x$ and $y$ have unequal coefficients $h$ and $k$ respectively in $f$. Let $H$ and $K$ be the components of $G:(E(G)-||g||)$ containing $x$ and $y$ respectively. They are distinct, by 9.11.

Let $f_H$ be the 0-chain in which the members of $V(H)$ have coefficient 1 and all other vertices of $G$ have coefficient 0. Then $\|\delta f_H\| \subseteq \|\delta f\| = \|g\|$ and therefore $\|\delta f_H\| = \|g\|$ since $g$ is elementary. We deduce that each edge of $\|g\|$ has one end in $V(H)$ and similarly one end in $V(K)$.

9.13 *A set $Y \subseteq E(G)$ is an atom of $M(N)$ if and only if there are components $H$ and $K$ of $G$: $(E(G)-Y)$ such that each edge of $Y$ has one end in $V(H)$ and one in $V(K)$.*

PROOF: If $Y \epsilon M(N)$ the above condition is satisfied, by 9.12.

Conversely suppose $G$: $(E(G)-Y)$ has components $H$ and $K$ with the stated property. Clearly $Y$ is the domain of a coboundary $\delta f$ such that the vertices of $H$ have coefficient 1 in $f$, and those of $V(G)-V(H)$ coefficient 0.

Suppose however that there is a coboundary $\delta f_1$ of $G$ such that $||\delta f_1||$ is a proper subset of $Y$. Then there is a component $J$ of $G$: $(E(G)-\|\delta f_1\|$ which contains both $H$ and $K$. Hence $\delta f_1 = 0$, by 9.11 and the definition of a coboundary.

Theorem 9.13 provides another proof that $M(N)$ is independent of $R$. Its atoms are the bonds of $G$.

If $Y$ is any bond of $G$ then the graphs $H$ and $K$ of 9.13 are the *end-graphs* of $Y$ in $G$.

If $S \subseteq E(G)$ the common vertices of $G \cdot S$ and $G \cdot (E(G)-S)$ are the *vertices of attachment* of $S$ (and of $E(G)-S$) in $G$. We denote their number by $u(S)$. They may also be defined as those vertices of $G$ which are incident with edges of both $S$ and $E(G)-S$. If $u(S) = 1$ then the single vertex of attachment of $S$ is called a *cut-vertex* of $G$.

Let $x$ be any vertex of $G$. The star $St(x)$ of $x$ in $G$ is the set of all edges having one end at $x$ and one in

$V(G)-x$. We also write $L(x)$ for the set of all edges (loops) of $G$ whose ends coincide at $x$.

The remaining theorems of this section relate the structure of $G$ to that of $B(G)$.

9.14 *Let $S$ be a subset of $E(G)$ such that $w(S) \leq 1$. Then $S$ is a separator of $B(G)$.*

PROOF: Let $Y$ be any bond of $G$, with end graphs $H$ and $K$. We may suppose $K$ to include no common vertex of $G \cdot S$ and $G \cdot (E(G)-S)$. Then the intersections of $K$ with $G \cdot S$ and $G \cdot (E(G)-S)$ have no common edge or vertex, and since $K$ is connected one of them is null. Hence either $V(K) \subseteq V(G \cdot S)$ or $V(K) \subseteq V(G \cdot (E(G)-S))$, whence it follows that either $Y \subseteq S$ or $Y \subseteq E(G)-S$. The theorem follows.

As a corollary we have

9.141 *If $a$ is a loop of $G$ then $\{a\}$ is a separator of $B(G)$.*

Such a separator may be called *trivial* since it meets no atom of $B(G)$.

9.15 *Let $x$ be any vertex of $G$ such that $St(x)$ is non-null, and let $G_1$ be the graph derived from it by deleting the loops of $L(x)$. Then $St(x)$ is a bond of $G$ if and only if $x$ is not a cut-vertex of $G_1$.*

PROOF: Let $G_2$ be the graph obtained from $G$ by deleting the vertex $x$ and its incident edges. Let those components of $G_2$ which include vertices incident with members of $St(x)$ in $G$ be enumerated as $H_1, H_2, \ldots, H_k$. Such components exist since $St(x)$ is non-null. (See fig. 9A.)

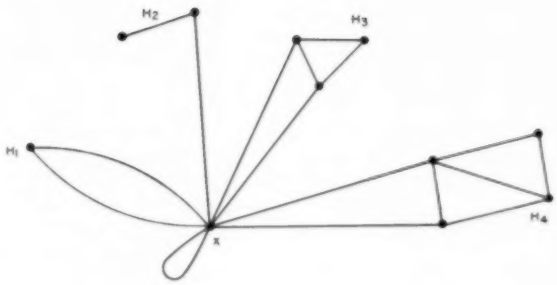



FIGURE 9A

Clearly $x$ is a cut-vertex of $G_1$ if and only if $k \geq 2$.

But the set of edges of $St(x)$ incident in $G$ with members of $V(H_i)$ is a bond of $G$ for each $i$, by 9.13, one of its end-graphs being $H_i$. Hence $St(x)$ is a bond of $G$ if $k=1$. But if $k \geq 2$ then $St(x)$ has a bond of $G$ as a proper subset, and is not itself an atom of $B(G)$, by Axiom I.

9.16 *$B(G)$ is a connected matroid if and only if $G \cdot E(G)$ is connected and $G$ has no cut-vertex.*

PROOF: If $G \cdot E(G)$ is not connected or if $G$ has a cut-vertex it is clear that some non-null proper subset $S$ of $E(G)$ satisfies $w(S) \leq 1$. Then $S$ is a separator of $B(G)$, by 9.14.

Conversely suppose $B(G)$ is not a connected matroid. Then there is a proper non-null subset $S$ of $E(G)$ which is a separator of $B(G)$.

Suppose $G$ has a loop $A$ incident with a vertex $x$. Then either $x$ is a cut-vertex of $G$, or $G \cdot \{A\}$ is a component of $G \cdot E(G)$. In either case the theorem is satisfied.

We may now assume that $G$ as no loop or cut-vertex. Let $x$ be any vertex of attachment of $S$. Then $St(x)$ is an atom of $B(G)$ meeting both $S$ and $E(G)-S$, by 9.15, which contradicts the definition of $S$ as a separator. We conclude that $w(S)=0$. Hence $G \cdot S$ is a union of components of $G \cdot E(G)$. This completes the proof of the theorem.

9.17 *Let $S$ be a non-null subset of $E(G)$. Then $S$ is an elementary separator of $B(G)$ if and only if the following conditions hold.*

(i) *$G \cdot S$ is connected and has no cut-vertex.*

(ii) *Each component of $G: (E(G)-S)$ has at most one vertex in common with $G \cdot S$.*

PROOF: Suppose first that $S$ is an elementary separator of $B(G)$. Then $B(G) \cdot S$ is a connected matroid, by 3.47. But $B(G) \cdot S = B(G \cdot S)$, by 3.321. Hence $G \cdot S$ is connected and has no cut-vertex, by 9.16.

It may happen that $S$ consists of a single loop of $G$. In that case proposition (ii) is trivially true. We may therefore assume that no edge of $S$ is a loop.

Let $x$ be any vertex of $G \cdot S$, and let $T$ be its star in $G \cdot S$. Then $T$ is an atom of $B(G \cdot S)$, of $B(G) \cdot S$, of $B(G) \times S$ and $B(G)$, by 3.321 and 3.42. Let $C(x)$ be its end-graph in $G$ which includes $x$. Clearly $C(x)$ is a component of $G: (E(G)-S)$, and $x$ is its only common vertex with $G \cdot S$.

Conversely suppose $S$ to satisfy (i) and (ii). If $C$ is any component of $G: (E(G)-S)$ then $w(E(C)) \leq 1$, by (ii), and therefore $E(C)$ is a separator of $B(G)$, by 9.14. The union of all such sets $E(C)$, namely $E(G)-S$, is thus also a separator of $B(G)$. Hence $S$ is a separator of $B(G)$.

If $S$ is not an elementary separator of $B(G)$ there is another non-null separator $S'$ of $B(G)$ such that $S' \subset S$. But then $S'$ is a separator of $B(G \cdot S)$, by 3.321 and 3.43. Hence (i) is contradicted, by 9.16.

This completes the proof.

Suppose we are given an elementary separator $S$ of $B(G)$. For each vertex $x$ of $G \cdot S$ we define $C(G, S, x)$ as that component of $G: (E(G)-S)$ which has $x$ as a vertex. If $x$ and $y$ are distinct vertices of $G \cdot S$ then $C(G, S, x)$ and $C(G, S, y)$ have no common vertex, by 9.17.

9.18 *Let $G$ be a connected graph, and let $S$ and $T$ be distinct elementary separators of $B(G)$. Then there are vertices $s$ and $t$, of $G \cdot S$ and $G \cdot T$, respectively, such that $G \cdot S$ is a subgraph of $C(G, T, t)$, and $G \cdot T$ is a subgraph of $C(G, S, s)$. Moreover each vertex of $G$ belongs to one or both of $C(G, T, s)$ and $C(G, T, t)$.*

PROOF: The first part of the theorem follows from the fact that $G \cdot S$ is a connected subgraph of $G: (E(G)-T)$, by 9.17, and from the analogous result for $G \cdot T$.

To prove the second part suppose $v$ is a vertex of $G$ not belonging to $C(G, S, s)$. Then, by the connection

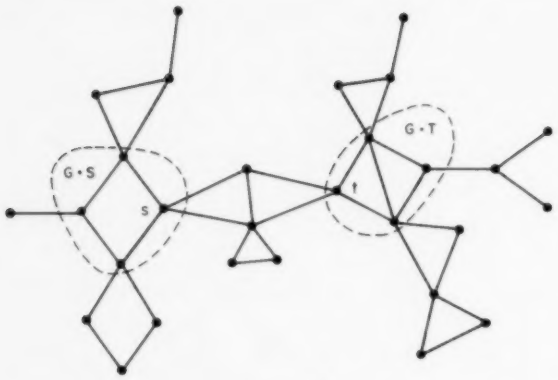



FIGURE 9B

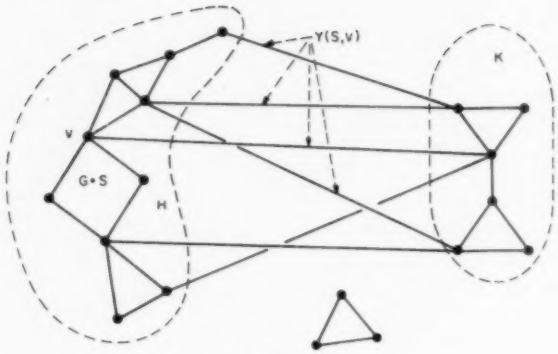



FIGURE 9C

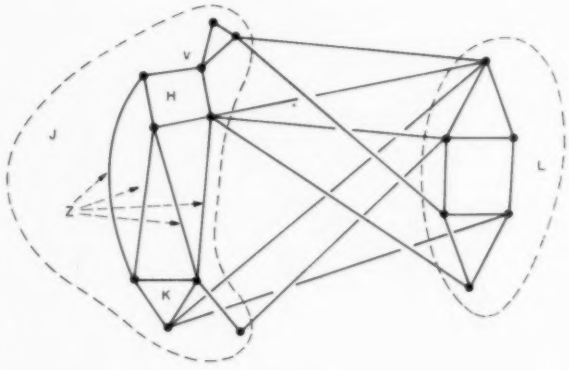



FIGURE 9D

of $G$, $v$ belongs to some $C(G, S, u)$ such that $u \neq s$ and $C(G, S, u)$ has no common vertex with $G \cdot T$. But then $c(G, S, u)$, with $G \cdot S$, is a subgraph of $C(G, T, t)$. (See fig. 9B.)

### 9.2. Bridges in a Bond-Matroid

Let $Y$ be a bond in a graph $G$, with end-graphs $H$ and $K$. Write $G_1 = G \cdot (E(G) - Y)$.

If $S$ is an elementary separator of $B(G_1)$, that is $B(G) \cdot (E(G) - Y)$, then the connected graph $G_1 \cdot S$ is a subgraph of some component of $G_1$. We refer to $G_1 \cdot S$, which is identical with $G \cdot S$, as a *bridge-graph* of $Y$ in $G$.

We note that any component of $G_1$ either consists solely of an isolated vertex, or it is a union of bridge-graphs of $Y$ in $G$. In particular this observation applies to $H$ and $K$. It leads immediately to the following important theorem.

9.21 *If $B(G) \cdot (E(G) - Y)$ is a connected matroid, then either $H$ or $K$ consists of a single vertex.*

This result is often helpful in determining whether a given binary matroid is graphic.

Suppose $G \cdot S$ is a bridge-graph of $Y$ in $G$. Let $v$ be one of its vertices. We write $Y(S, v)$ for the set of all edges of $Y$ having one end in $C(G_1, S, v)$. The set $Y(S, v)$ can thus be non-null only if $G \cdot S$ is contained in an end-graph of $Y$ in $G$. (See fig. 9C.)

9.22 *Let $S$ be a bridge of $Y$ in $B(G)$, so that $G \cdot S$ is a bridge-graph of $Y$ in $G$. Then either $\pi(B(G), S, Y)$ is $\{Y, \phi\}$ or it is the class of all non-null subsets $Y(S, v)$ of $Y$ such that $v \in V(G \cdot S)$.*

PROOF: Let $Z$ be any atom of $B(G) \cdot (E(G) - Y)$ on $S$. Then $Z$ is a bond of $G \cdot S$. Let its end-graphs in $G \cdot S$ be $H$ and $K$. These two subgraphs of $G$, together with the edges of $Z$ constitute the whole of $G \cdot S$, by 9.17.

Let $H_1$ be the union of $H$ with all the graphs $C(G_1, S, v)$ such that $v \in V(H)$, and let $K_1$ be defined similarly.

It may happen that one of $H_1$ and $K_1$ has no vertex incident in $G$ with a member of $Y$. In that case $Z$ is a bond of $G$ with either $H_1$ or $K_1$ as one end-graph. It then determines the trivial partition $\{Y, \phi\}$ of $Y$.

In the remaining case each of $H_1$ and $K_1$ has a vertex incident with a member of $Y$. Then $G \cdot S$ is contained in an end-graph $J$ of $Y$ in $G$. Let the other end-graph of $Y$ in $G$ be $L$. (See fig. 9D.) Then $H_1$ is an end-graph of a bond $Y_1$ of $G$, whose other end-graph is made up of $L$, $K_1$ and the edges of $Y$ joining $L$ to $K_1$. We have

$$Y_1 = Z \cup \left[ \bigcup_{v \in V(H)} Y(S, v) \right]$$

Since this argument applies for each choice of $Z$ we deduce that each member of $\pi(B(G), S, Y)$ is a union of sets of the form $Y(S, v)$. The argument assumes $S$ to be a nontrivial bridge, but the result is still trivially true in the remaining case.

Consider any vertex $v$ of $G \cdot S$. If $S$ consists of a single loop of $G$ we have $Y = Y(S, v)$. In the remaining case $S$ includes no loop of $G$, by 9.17. Let $Z$ be the star of $v$ in $G \cdot S$. Then $Z$ is a bond of $G \cdot S$ one of whose end-graphs consists solely of the vertex $v$, by 9.15

and 9.17. It therefore determines a partition of $Y$ having $Y(S, v)$ as one of its primary segments, by the preceding result.

Combining the above results we find that $Y(S, v)$ is either null or a member of $\pi(B(G), S, Y)$, for each $v \in V(G \cdot S)$, and hence that the theorem is true.

This theorem may be regarded as verifying 8.62 for graphic matroids.

9.23 *Every graphic matroid is even.*

PROOF: Let $M$ be a graphic matroid. Then $M = B(G)$ for some graph $G$. Let $Y$ be any atom of $M$, that is any bond of $G$, and let the end-graphs of $Y$ in $G$ be $H$ and $K$.

Let $S_1$ and $S_2$ be overlapping bridges of $Y$ in $B(G)$. The bridge graphs $G \cdot S_1$ and $G \cdot S_2$ are each contained in an end-graph of $G$.

Suppose they are subgraphs of the same end-graph $H$. There are vertices $v_1$ of $G \cdot S_1$ and $v_2$ of $G \cdot S_2$ such that $G \cdot S_2$ is a subgraph of $C(G_1, S_1, v)$ and $G \cdot S_1$ is a subgraph of $C(G_2, S_2, v)$, by 9.18, applied to $H$. Moreover

$$Y = Y(S_1, v_1) \cup Y(S_2, v_2)$$

by the same theorem. This contradicts the hypothesis that $S_1$ and $S_2$ overlap, by 9.22.

We deduce that if $S_1$ and $S_2$ overlap then $G \cdot S_1$ and $G \cdot S_2$ are subgraphs of distinct end-graphs of $G$.

Let $P$ be the class of all bridges $S$ of $Y$ in $M$ such that $G \cdot S$ is a subgraph of $H$, and let $Q$ be the class of all other bridges of $Y$ in $M$. Then, by the result just proved no members of the same class $P$ or $Q$ overlap. Thus $Y$ is bridge-separable.

It follows that $M$ is an even matroid.

Our next theorem shows that two distinct graphs, even though without cut-vertices, can have identical bond-matroids.

Let $G$ be a graph and let $S$ be a subset of $E(G)$ such that $w(S) = 2$. Let the vertices of attachment of $S$ be $x$ and $y$. We construct a graph $G'$ such that $E(G) = E(G')$, $V(G) = V(G')$, and the same incidence relations hold in $G'$ as in $G$, with the following exception. If $A \in S$, then $A$ is incident with $x$ or $y$ in $G'$ if and only if it is incident with $y$ or $x$ respectively in $G$. We say $G'$ is obtained from $G$ by *reversing* $S$. We can recover $G$ from $G'$ by reversing $S$ again. We illustrate the operation in figure 9E.

9.24 *$B(G)$ is invariant under the operation of reversing a subset $S$ of $G$ such that $w(S) = 2$.*

PROOF: Let $g$ be the 0-chain of $G$ over $GF(2)$ such that $g(z) = 1$ when $z$ is a vertex of $G \cdot S$ other than $x$ or $y$, and $g(z) = 0$ in the remaining cases.

If $f$ is any 0-chain on $G$ over $GF(2)$ it is readily verified that the coboundary of $f$ in $G$ is the coboundary of $f + g$ in $G'$. Hence $G$ and $G'$ have the same coboundary group over $GF(2)$. The theorem follows, by the definition of a bond-matroid.

9.25 *Let $Y$ be a totally bridge-separable atom of a connected graphic matroid $M$. Then there is a connected graph $G$, having no cut-vertex, such that $M = B(G)$ and some end-graph of $Y$ in $G$ consists of a single vertex.*

PROOF: There exists $G$ such that $M = B(G)$. We may suppose $G$ to be without isolated vertices, since $M = B(G) \cdot E(G) = B(G \cdot E(G))$. Then $G$ is connected and has no cut-vertex, by 9.17. We write $G_1 = G \cdot (E(G) - Y)$ as before.

Let the end-graphs of $Y$ in $G$ be $H_1$ and $H_2$. By the connection of $G$ these are the only components of $G_1$. We suppose $G$, $H_1$, and $H_2$ chosen so that $|E(H_2)|$ has the least possible value.

Assume $|E(H_2)| > 0$. Then also $|E(H_1)| > 0$, by the above choice.

Choose bridge-graphs $G \cdot S_1$ and $G \cdot S_2$ of $Y$ in $G$ contained in $H_1$ and $H_2$ respectively. Since the bridges $S_1$ and $S_2$ do not overlap there are, by 9.22, vertices $v_1$ and $v_2$ of $G \cdot S_1$ and $G \cdot S_2$ respectively such that

(i) $$Y(S_1, v_1) \cup Y(S_2, v_2) = Y.$$

Keeping $S_2$ and $v_2$ fixed we consider all possible choices of $S_1$ and $v_1$ for which (i) is true, and we select one for which $C(G_1, S_1, v_1)$ has the least number of edges.

Consider the set of all bridges $S$ of $Y$ in $M$ such that $G \cdot S$ has $v_1$ as a vertex. (See fig. 9F.) Let them be enumerated as $P_1, P_2, \ldots, P_k$, taking $P_1 = S_1$.

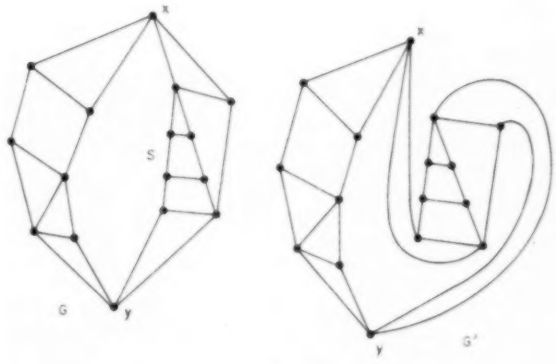



FIGURE 9E

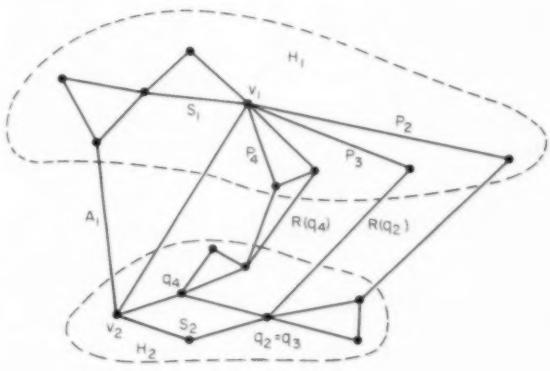



FIGURE 9F

For each $P_j$ let $F_j$ be the subgraph of $H_1$ which is the union of $G \cdot P_j$ and those subgraphs $C(G_1, P_j, x)$ of $G_1$ for which $x \epsilon V(G \cdot P_j) - \{v_1\}$. The graphs $F_j$ have the common vertex $v_1$, and each of them is connected. By 9.17 no two of them have any common vertex other than $v_1$. Since $H_1$ is connected it is the union of the graphs $F_j$.

For each $P_j$ there is a vertex $p_j$ of $G \cdot P_j$ and a vertex $q_j$ of $G \cdot S_2$ such that

(ii) $$Y(P_j, p_j) \cup Y(S_2, q_j) = Y.$$

This follows from the hypothesis that $Y$ is totally bridge-separable, and from 9.22. In accordance with (i) we take $p_1 = v_1$ and $q_1 = v_2$.

Since $G$ has no cut-vertex it follows that for each $P_j$ there is an edge $A_j$ in $Y(S_2, q_j)$ but not in $Y(P_j, p_j)$, and an edge $D_j$ in $Y(P_j, p_j)$ but not in $Y(S_2, q_j)$.

Suppose $p_j \neq v_1$. Considering the edge $A_1$, of which one end is a vertex of $F_1$ other than $v_1$, we find that $q_j = q_1 = v_2$, by (ii). But in this case $C(G_1, P_j, p_j)$ is a subgraph of $C(G, S_1, v_1)$, and has fewer edges than $C(G_1, S_1, v_1)$. This however contradicts the definition of $S_1$ and $v_1$. We deduce that in fact

(iii) $$p_j = v_1$$

for each $j$.

Considering the edge $A_j$ we see that $q_j$ is uniquely determined for each $p_j$. Let $Z_j$ be the set of all edges of $Y$ having one end a vertex of $F_j$ other than $v_1$. Then $Z_j$ is non-null since it includes $A_j$. By (ii) each edge of $Z_j$ has one end a vertex of $C(G_1, S_2, q_j)$.

For each vertex $x$ of $G \cdot S_2$ we define $R(x)$ as the subgraph of $G$ formed by taking the union of $C(G_1, S_2, x)$ and those graphs $F_j$ for which $q_j = x$, and then adjoining the members of the corresponding sets $Z_j$ as new edges.

For a given $x$ the graph $R(x)$ may have only the one vertex $x$. If this is not so the set $E(R(x))$ is non-null and its vertices of attachment in $G$ are $x$ and $v_1$, since $x$ is not a cut-vertex of $G$.

Let us reverse all the non-null sets $R(x)$ (which are disjoint). By 9.24 $G$ is then transformed into another graph $G'$ with the same bond-matroid $M$. We note that $G'$ has no isolated vertex. But the transformation replaces $H_1$ by an end-graph $K$ of $Y$ in $G'$ such that $E(K) = E(H_2) - S_2$. But this is contrary to the choice of $G$, $H_1$, and $H_2$.

We deduce that in fact $|E(H_2)| = 0$. The theorem follows.

## 9.3. Property of Odd Matroids

In this section we prove the following theorem.

9.31 *Every odd binary matroid has a minor of Type BII, KI or KII.*

PROOF: Assume that the theorem fails. Then there exists an odd binary matroid $M$, on a set $E$, which has no minor of Type BII, KI, or KII. Choose such a matroid so that $|E|$ has the least possible value.

By definition there is at least one atom $Y$ of $M$ which is not bridge-separable.

There must be a cyclic sequence

$$P = (B_1, B_2, \ldots, B_{2n+1}, B_1)$$

of an odd number $2n+1$, where $n \geq 1$, of distinct bridges $B_i$ of $Y$ in $M$ such that each overlaps its two neighbors in the sequence. We refer to $P$ as an *odd overlap-circuit* of $Y$.

We observe here another intrusion of graph theory into the general theory of matroids. The bridges $B_i$ can be regarded as the vertices of a graph whose edges are the overlapping pairs. We have used the well-known theorem that a graph is bipartite if and only if it has no polygon with an odd number of edges.

We proceed by a sequence of subsidiary theorems.

9.311 *Let $Y$ be any atom of $M$ which is not bridge-separable. Then the bridges of $Y$ in $M$ can be arranged in a single odd overlap-circuit $P$ so that two bridges overlap only if they are consecutive in $P$.*

PROOF: Choose an odd overlap-circuit $P$ of $Y$ having as few terms as possible. This clearly implies that nonconsecutive members of $P$ do not overlap.

Let $S$ be the union of the members of $P$. Then $P$ is an odd overlap-circuit of $M \times (S \cup Y)$, by 8.53. But every minor of $M \times (S \cup Y)$ is a minor of $M$. Hence $E = S \cup Y$, by the choice of $M$. The theorem follows.

In the next few propositions we suppose given an atom $Y$ of $M$, with a corresponding odd overlap-circuit

$$P = (B_1, B_2, \ldots, B_{2n+1}, B_1) \cdot$$

9.312 *Suppose there is an atom $Z$ of $M \cdot (E - Y)$ on $B_i$ which is skew to both $B_{i-1}$ and $B_{i+1}$. Then $B_i$ consists of a single cell.*

PROOF: With regard to the suffices we adopt the convention that $B_0 = B_{2n+1}$, $B_{-1} = B_{2n}$, and in general $B_r = B_{2n+1+r}$.

Write $S = E - (B_i - Z)$. The bridges $B_j$ of $Y$ other than $B_i$ are also bridges of $Y$ in $M \times S$, determining the same partitions of $Y$ as in $M$, by 8.53.

On the other hand $Z$ is an atom of $((M \cdot (E - Y)) \times B_i) \times Z$, that is of $(M \cdot B_i) \times Z$ by 3.332 and 3.42. But this matroid is $(M \times S) \cdot Z = ((M \times S) \cdot (S - Y)) \cdot Z$. Hence $Z$ is the only remaining bridge of $Y$ in $M \times S$.

Since $Y \cup Z$ is a line of $M$ on $M \times S$ the atom $Z$ determines the same partition $\{U, V\}$ of $Y$ in $M \times S$ as in $M$. Each of the sets $U$ and $V$ meets at least two members of $\pi(M, B_{i-1}, Y)$ and at least two of $\pi(M, B_{i+1}, Y)$.

From the above results we deduce that the replacement of $B_i$ by $Z$ in $P$ yields an odd overlap-circuit of $M \times S$. Hence $B_i$ is monatomic, with the single atom $Z$, by the choice of $M$.

Now let $a$ be any cell of $Z$. Write $T = E - (Z - \{a\})$. We have $Y \epsilon M \cdot T$, by 8.531. The bridges of $Y$ in $M$, other than $Z$, are also bridges of $Y$ in $M \cdot T$, determining the same partitions of $Y$ as in $M$, by 8.531. The only other bridge of $Y$ in $M \cdot T$ is the atom $\{a\}$.

$Z$ determines the partition $\{U, V\}$ of $Y$ in $M$. There exists $X' \epsilon M \cdot T$ such that $\{a\} \subseteq X' \subseteq Z \cup U \epsilon M$. There

exists $X \in M$ such that $X \cap T = X'$. But this implies $X \subseteq Z \cup U$, and therefore $X = Z \cup U$, by Axiom I. Hence $X' = U \cup \{a\}$. Accordingly $a$ determines the same partition of $Y$ in $M \cdot T$ as does $Z$ in $M$.

We deduce that the replacement of $B_i = Z$ by $\{a\}$ in $P$ yields an odd overlap-circuit of $M \cdot T$. Hence $B_i = \{a\}$, by the choice of $M$.

We refer to a bridge consisting of a single cell as *unicellular*.

9.313 *If $n \geq 2$, then every bridge of $Y$ in $M$ is unicellular*

PROOF: Consider any bridge $B_i$ in $P$. Now $B_i$ and $B_{i+1}$ are not equipartite 3-bridges, for otherwise $B_{i-1}$ and $B_{i+1}$ would overlap, contrary to 9.311. Hence $B_i$ is skew to $B_{i+1}$, and similarly to $B_{i-1}$, by 8.44. But $B_{i-1}$ and $B_{i+1}$ are not skew, by 9.311. Hence there is an atom $Z$ of $M \cdot (E - Y)$ on $B_i$ which is skew to both $B_{i-1}$ and $B_{i+1}$, by 8.42. Accordingly $B_i$ is unicellular, by 9.312.

Let $\Sigma$ be the class of all non-null subsets of $Y$ of the form

$$\bigcap_{i=1}^{2n=1} W_i,$$

where $W_i \in \pi(M, B_i, Y)$. Clearly the members of $\Sigma$ are disjoint and their union is $Y$.

9.314 *If $Q \in \Sigma$, then $|Q| = 1$. Moreover there is an integer $i$ such that $Y - Q = W_i \cup W_{i+1}$, where $W_i \in \pi(M, B_i, Y)$ and $W_{i+1} \in \pi(M, B_{i+1}, Y)$.*

PROOF: Choose $a \in Q$. Now $Y - \{a\} \in M \cdot (E - \{a\})$, by 8.54. But $P$ is not an odd overlap-circuit of $Y - \{a\}$ in $M \cdot (E - \{a\})$, by the choice of $M$. It follows from 8.54 that $Q - \{a\}$ must be null. Moreover there is an integer $i$ such that $B_i$ and $B_{i+1}$ do not overlap as bridges of $Y - \{a\}$ in $M \cdot (E - \{a\})$. The proposition follows.

Suppose $n \geq 2$. Then each bridge of $Y$ in $M$ is unicellular, by 9.313. Write $B_i = \{b_i\}$. For each $i$ the atom $\{b_i\}$ of $M \cdot (E - Y)$ determines a partition

$$\pi(M, B_i, Y) = \{S_i, T_i\}$$

of $Y$. Since consecutive members of $P$ overlap neither $S_i$ nor $T_i$ is null.

Consider a particular bridge $B_i$. By 9.311 we can adjust the notation so that $S_j \subset S_i$ or $S_j \subset T_i$ whenever $B_j$ is distinct from and not consecutive with $B_i$ in $P$. We choose one $B_j$ of this kind and arrange, by interchanging $S_i$ and $T_i$ if necessary, that $S_j \subset T_i$.

Let $B_k$ be the first member of the sequence

$$(B_j, B_{j+1}, \ldots, B_{i-2})$$

such that $S_{k+1} \cap S_i$ is non-null. Then $S_{k+1}$ meets both $S_i$ and $T_i$, since $B_{k+1}$ is skew to $B_k$. Hence $B_{k+1}$ is consecutive with $B_i$ in $P$, that is $B_k = B_{i-2}$.

Similarly if $B_l$ is the last member of the sequence

$$(B_{i+2}, B_{i+3}, \ldots, B_j)$$

such that $S_{l-1} \cap S_i$ is non-null we find that $l = i + 2$.

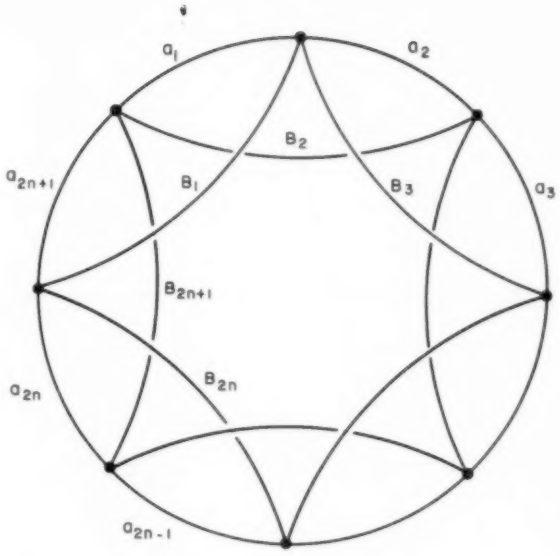



FIGURE 9G

It follows that $S_h \subset T_i$ for each $B_h$ not consecutive with $B_i$ in $P$. $(h \neq i.)$

By the result just proved we can adjust the notation so that $S_i$ contains neither $S_j$ nor $T_j$ as a subset whenever $B_i$ and $B_j$ are distinct. Then by 9.311 we find that $S_i \cap S_j$ is non-null if and only if $B_i$ and $B_j$ are consecutive in $P$.

It follows that the $2n+1$ sets $S_i \cap S_{i+1}$, $(1 \leq i \leq 2n+1)$ are distinct members of $\Sigma$. We must now show that they are the only members of $\Sigma$.

Choose any $W \in \Sigma$. By 9.314 we can write $W = S_i \cap S_{i+1}$, $S_i \cap T_{i+1}$, $T_i \cap S_{i+1}$ or $T_i \cap T_{i+1}$, for some $i$. In the first case there is nothing to prove. In the second case we observe that $S_{i-1}$ meets $S_i$ and is a subset of $T_{i+1}$. Hence $W$ is the member $S_{i-1} \cap S_i$ of $\Sigma$. We deal with the third case similarly, using $S_{i+2}$ instead of $S_{i-1}$. In the fourth case we have $S_{i+3} \subseteq T_i \cap T_{i+1} = W$, whence $S_{i+3} = W$. But then $B_{i+3}$ overlaps no other bridge of $Y$, since $T_{i+3} \cup W = Y$, and this is a contradiction.

We are now in a position to prove

9.315 *The assumption that $n \geq 2$ leads to the contradiction that $M$ is a matroid of Type KI.*

PROOF: The above diagram represents $M$, on the as yet unproved assumption that it is cographic. We write $a_i$ for the single cell of the member $S_i \cap S_{i+1}$ of $W$. (See 9.314.)

If $n > 2$ we write $Y' = (Y - S_1) \cup B_1$. By 8.55 $Y' \in M$ and $B_3, B_4, \ldots, B_{2n}$ are bridges of $Y'$ in $M$. Moreover there is another bridge $B'$ which contains $S_1$, $B_2$, and $B_{2n+1}$. It also follows from 8.55 that $(B', B_3, B_4, \ldots, B_{2n}, B')$ is an odd overlap-circuit of $Y'$ in $M$. This contradicts 9.312 since $B'$ is not unicellular.

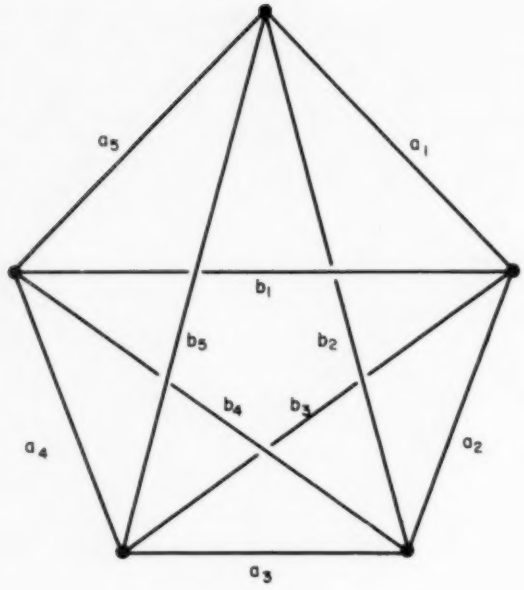



FIGURE 9H

In the remaining case $n=2$. (Fig. 9H.)

Let us write $a_i=L_{i,\,i+1}$ if $1\leqslant i\leqslant 4$, and $a_5=L_{1,\,5}$. We put also $b_i=L_{i,\,i+2}$ if $1\leqslant i\leqslant 3$, $b_4=L_{1,\,4}$, and $b_5=L_{2,\,5}$. We construct a graph $G$ such that $E(G)=E$. We take it to have just five vertices $v_1$, $v_2$, $v_3$, $v_4$, and $v_5$, the ends of $L_{i,\,j}$ in $G$ being $v_i$ and $v_j$. Then $G$ is a complete 5-graph.

We find that $M=P(G)$. To prove this we observe that $Y$ and the five sets $\{b_i\}\cup S_i$ are atoms of both $M$ and $P(G)$. Any atom $X$ of $M$ is a linear combination of these six. For otherwise some combination of atoms of $M$ would be a non-null proper subset of $Y$, which, by 5.34, is contrary to Axiom I. Hence $X$ is a mod 2 sum of atoms of $P(G)$ and so has an atom $X'$ of $P(G)$ as a subset, by 5.34. Similarly any atom of $P(G)$ contains an atom of $M$.

We deduce that $M=P(G)$, by Axiom I. Thus $M$ is of Type KI.

We may now assume that no atom of $M$ has an odd overlap-circuit constituted by more than three bridges.

We select a particular atom $Y$ which is not bridge-separable, and we denote its three mutually overlaping bridges by $B_1$, $B_2$, and $B_3$.

In our analysis we make use of the following proposition.

9.316 *Suppose $W_1$ is a member of $\pi(M, B_1, Y)$ with the property that some atom $Z$ of $M\cdot(E-Y)$ on $B_1$ determines the partition $\{W_1, Y-W_1\}$ of $Y$. Then either $\pi(M, B_2, Y)$ or $\pi(M, B_3, Y)$ has a member which cuts $W_1$.*

PROOF: Assume the contrary. Then there exist $W_2\epsilon\pi(M, B_2, Y)$ and $W_3\epsilon\pi(M, B_3, Y)$ such that $W_1\subseteq W_2\cap W_3$.

Let $Y'$ denote the atom $(Y-W_1)\cup Z$ of $M$. It follows from 8.55 that $B_2$ and $B_3$ are overlapping bridges of $Y'$ in $M$.

Now $W_1$ is an atom of $M\cdot(E-Y')$, by 8.55. Suppose it meets another atom $X$ of $M\cdot(E-Y')$. Since $X-W_1$ is non-null, and $X$ cannot be an atom of the bridge $B_2$ or $B_3$ of $Y'$ we have $X\subseteq B_1\cup Y'\subset B_1\cup Y$. There is thus an atom of $M\times(B_1\cup Y)$ which cuts $W_1$. This is impossible, by 8.51 and 8.52. We deduce that $W_1$ is another bridge of $Y'$ in $M$. Determining the partition $\{Y-W_1, Z\}$ of $Y'$ it overlaps neither $B_2$ nor $B_3$.

Let $H$ be the class of all remaining bridges of $Y'$ in $M$. The union of the members of $H$ is $B_1-Z$. This set is non-null since $Z$ determines the partition $\{Y-W_1, W_1\}$ of $Y$, while $B_1$ overlaps $B_2$ and $B_3$.

Applying 8.55 with $Y$ and $Y'$ interchanged we see that if $B\epsilon H$, then no member of $\pi(M, B, Y')$ can contain $Z$.

Since $B_1$ and $B_2$ overlap with respect to $Y$ there is an atom $X$ of $M$ on $B_1\cup Y-W_1=B_1\cup Y'$ which cuts $Y-W_2$. Then $X\cap(E-Y')$ is a union of disjoint atoms of $M\cdot(E-Y')$, all contained in $B_1\cup Y'$, by 8.51. Let us enumerate them as $X_1, X_2, \ldots, X_k$. Let $X_i$ determine the partition $\{S_i, T_i\}$ of $Y'$. The mod 2 sum of $X$ and the atoms $X_i\cup S_i$ of $M$ is contained in $Y'$, and is either $Y'$ or $\phi$. by 5.32 and Axiom I. Hence some $X_i$ cuts $Y-W_2$. This $X_i$ must be on some member of $H$.

We deduce that there exists $C_1\epsilon H$ such that no member of $\pi(M, C_1, Y')$ contains either $Z$ or $Y-W_2$. Then $C_1$ overlaps both $W_1$ and $B_2$ as bridges of $Y'$. Similarly there exists $C_2\epsilon H$ overlapping both $W_1$ and $B_3$.

We may have $C_1=C_2$. In that case $(B_2, B_3, C_1)$ is an odd overlap-circuit of $Y'$ which does not include $W_1$, and this contradicts 9.311.

In the remaining case $Y'$ has an odd overlap-circuit $(B_2, B_3, C_2, W_1, C_1, B_2)$ having five members. But such odd overlap-circuits are ruled out by 9.315.

The proposition follows.

By 8.43 and 9.316 $B_1$, $B_2$, and $B_3$ are not equipartite 3-bridges.

Suppose $\pi(M, B_1, Y)=\pi(M, B_2, Y)=\{W_1, W_2, W_3\}$. Then $B_3$ is skew to $B_1$ and $B_2$, by 8.44. But $B_1$ and $B_2$ are clearly not skew to each other. Hence $B_3$ is unicellular, by 9.312. It determines a partition $\{S, T\}$ of $Y$.

We may suppose without loss of generality that $W_1\cap S$ is non-null. Then $W_1\cap S\epsilon\Sigma$. Moreover $W_1\cap S$ is a proper subset of $W_1$, by 8.43 and 9.316. In view of 9.314 we may write, without loss of generality,

$$Y-(W_1\cap S)=T\cup W_i,$$

where $i=1$ or 2. In the apparent alternative $Y-(W_1\cap S)=S\cup W_i$ we have $Y=S\cup W_i$, which is impossible since $B_1$ and $B_3$ overlap. But then $W_3\subseteq T$, $W_3\epsilon\Sigma$, and we have a contradiction of 8.43 and 9.316.

We may now suppose that $B_1$, $B_2$, and $B_3$ are mutually skew, by 8.44. Applying 8.42 and 9.312 we find that two of them, say $B_1$ and $B_2$, are unicellular.

Let them determine partitions $\{S_1, T_1\}$ and $\{S_2, T_2\}$ of $Y$ respectively. We write

$$\pi(M, B_3, Y) = \{W_1, W_2, \ldots, W_k\}.$$

The number $k$ is at least 2, since the three bridges overlap.

9.317 *The assumption that $k=2$ leads to the contradiction that $M$ has a minor of Type BII or KII.*

PROOF: On this assumption $B_3$ must be unicellular, by 9.312.

Suppose $W_1$ does not meet both $S_1 \cap S_2$ and $T_1 \cap T_2$. Then it meets both $S_1 \cap T_2$ and $S_2 \cap T_1$, for it is not contained in any of the sets $S_1$, $S_2$, $T_1$, $T_2$, by the overlapping of the bridges. We may therefore adjust the notation so that $W_1$ meets both $S_1 \cap S_2$ and $T_1 \cap T_2$.

Assume that $W_1$ contains neither $S_1 \cap T_2$ not $S_2 \cap T_1$. Then $M$ has a minor of Type BII, by 8.64, which is contrary to the choice of $M$. We may therefore make a further adjustment of notation so that $S_1 \cap T_2 \subseteq W_1$.

We now observe that $W_1$ cuts both $S_1 \cap S_2$ and $T_1 \cap T_2$. For otherwise $W_2$ would be a subset of $T_1$ or $S_2$, and $B_3$ would not overlap both $B_1$ and $B_2$. Another application of 8.64 shows that $S_2 \cap T_1 \subseteq W_2$.

We can now list the members of $\Sigma$, each unicellular by 9.314, as follows.

$$S_1 \cap S_2 \cap W_1 = \{L_{1,1}\},$$

$$S_1 \cap S_2 \cap W_2 = \{L_{2,1}\}$$

$$S_2 \cap T_1 \cap W_2 = \{L_{2,2}\}$$

$$T_1 \cap T_2 \cap W_2 = \{L_{3,2}\}$$

$$T_1 \cap T_2 \cap W_1 = \{L_{3,3}\}$$

$$S_1 \cap T_2 \cap W_1 = \{L_{1,3}\}.$$

We also write $B_1 = \{L_{2,3}\}$, $B_2 = \{L_{1,2}\}$ and $B_3 = \{L_{3,1}\}$. This situation is represented in figure 9I, on the assumption that $M$ is cographic.

We now construct a graph $G$ with six vertices $a_1$, $a_2$, $a_3$, $b_1$, $b_2$, $b_3$, taking the cells $L_{i,j}$ as edges. The ends of $L_{i,j}$ are $a_i$ and $b_j$. Thus $G$ is a Thomsen graph. By an argument like the one which concludes the proof of 9.315 we find that $M = P(G)$.

Thus $M$ is a matroid of Type KII, contrary to the choice of $M$.

We may now suppose $k \geq 3$.

One of $S_1 \cap S_2$ and $T_1 \cap T_2$ must be contained in a member of $\pi(M, B_3, Y)$. For otherwise we could find an atom $Z$ of $M \cdot (E-Y)$ on $B_3$ cutting both $S_1 \cap S_2$ and $T_1 \cap T_2$, by 4.37 and 8.41. Then $Z$ would be skew to both $B_1$ and $B_2$, and $B_3$ would be unicellular, by 9.312. This is impossible since $k \geq 3$. Similarly one of $S_1 \cap T_2$ and $S_2 \cap T_1$ is contained in a member of $\pi(M, B_3, Y)$.

We may thus adjust the notation so that

$$S_1 \cap S_2 \subseteq W_1, \qquad S_1 \cap T_2 \subseteq W_2.$$

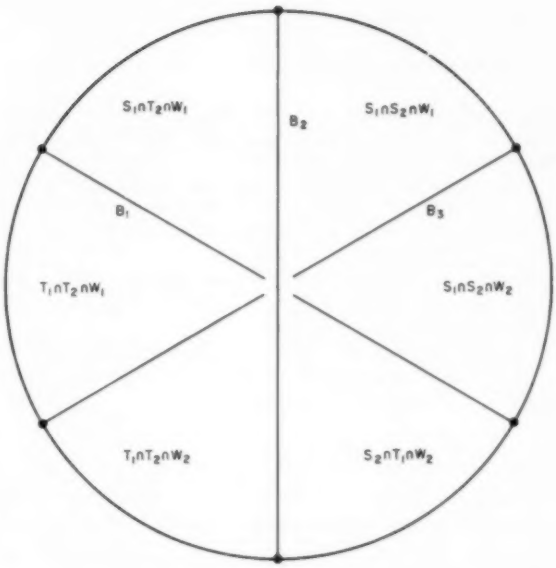



FIGURE 9I

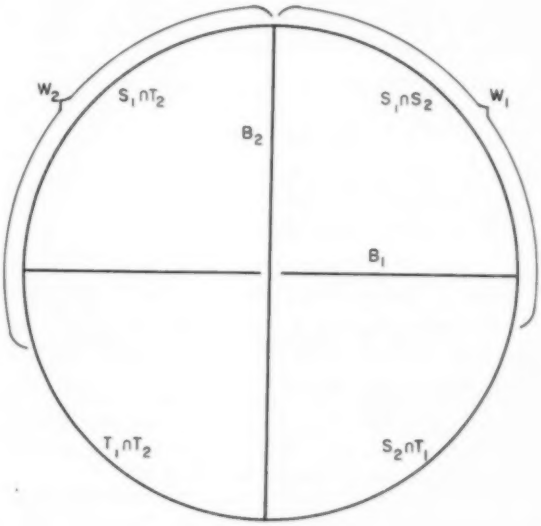



FIGURE 9J

For $S_1 \cap S_2$ and $S_1 \cap T_2$ are not both in $W_1$: otherwise $B_1$ and $B_3$ would not overlap.

We note that $W_1$ does not meet $T_1 \cap T_2$. For otherwise $W_1$ would contain $S_1 \cap T_2$ or $S_2 \cap T_1$, by 8.64, since $M$ has no minor of Type BII. But then $W_1$ would contain $S_1$ or $T_2$, and $B_3$ would not overlap both $B_1$ and $B_2$. Similarly $W_2$ does not meet $S_2 \cap T_1$.

Suppose $W_1$ meets $S_2 \cap T_1$. This case is represented schematically in figure 9J.

By 9.314 we have

$$Y-(W_1\cap S_2\cap T_1)=S_1\cup T_2, S_1\cup W_x \text{ or } T_2\cup W_x,$$

for a suitable suffix $x$. The first alternative implies that $S_2\cap T_1\subseteq W_1$. But then $S_2\subseteq W_1$ which is impossible since $B_2$ and $B_3$ overlap.

The second possibility requires $T_1\cap T_2\subseteq W_x$ and $W_1\cup W_2\cup W_x=Y$. Thus $x=k=3$. But now $W_2$ is a subset of $S_1$ and of $T_2$. Hence this case is ruled out by 8.43 and 9.316.

In the third case $S_1\cap S_2\subseteq W_x$, so that $W_x=W_1$. But it is then necessary that $S_2\cap T_1\subseteq W_1$. This is impossible since it requires $S_2\subseteq W_1$.

We deduce that $W_1=S_1\cap S_2$. Similarly $W_2=S_1\cap T_2$.

Let $W_3$ be a member of $\pi(M, B_3, Y)$ which meets $S_2\cap T_1$. Then by 9.314 we have

$$Y-(W_3\cap S_2\cap T_1)=S_1\cup T_2, S_1\cup W_x \text{ or } T_2\cup W_x.$$

The first alternative implies $S_2\cap T_1\subseteq W_3$, the second that $k=3$ or 4 and $T_1\cap T_2\subseteq W_x$. The third implies $S_1\cap S_2\subseteq W_x=W_1$. Then $S_2\cap T_1\subseteq W_3$.

Similar results are obtained when we consider a member of $\pi(M, B_3, Y)$ meeting $T_1\cap T_2$. Since $S_2\cap T_1$ and $T_1\cap T_2$ are not contained in the same member of $\pi(M, B_3, Y)$, by the overlapping of $B_1$ and $B_3$, it follows that we can adjust the notation so that

$$\pi(M, B_3, Y)=\{W_1, W_2, W_3, W_4\},$$

$$W_1=S_1\cap S_2,$$

$$W_2=S_1\cap T_2,$$

$$W_3\subseteq S_2\cap T_1,$$

$$T_1\cap T_2\subseteq W_4.$$

Now there is an atom $Z$ of $M\cdot(E-Y)$ on $B_3$ which cuts both $W_1\cup W_2$ and $W_1\cup W_3$, by 4.37 and 8.41. This atom must determine one of the partitions $\{W_1, Y-W_1\}$ and $\{W_1\cup W_4, W_2\cup W_3\}$ of $Y$. The latter requires $M$ to have a minor of Type BII, by 8.64, and can be ruled out. But now we have a contradiction of 9.316, since $W_1\subseteq S_1$ and $W_1\subseteq S_2$.

The analysis is now complete. No matroid satsifies the conditions imposed on $M$, and theorem 9.31 is established.

## 9.4. Characterization of Graphic Matroids

One more major result is required to complete our theory.

9.41 *If a matroid $M$ is regular and has no minor of Type KI or KII, then $M$ is graphic.*

PROOF: Assume the contrary. Then there is a regular matroid $M$ on a set $E$ which has no minor of Type KI or KII but which is not graphic. Choose $M$ so the $dM$ has the least possible value.

Suppose first that $M$ is not connected. We can find complementary non-null separators $S_1$ and $S_2$ of $M$. There exist graphs $G_1$ and $G_2$ such that

$$B(G_1)=M\times S_1,\ B(G_2)=M\times S_2,$$

by 5.62. We may take $V(G_1)$ and $V(G_2)$ to be disjoint subsets of some larger set $V$. Then $G_1$ and $G_2$ together constitute a graph $G$. We note that $S_1$ and $S_2$ are separators of $B(G)$, by 9.14.

If $X\epsilon B(G)$ we have

$$X\epsilon B(G)\times S_i=B(G)\cdot S_i=B(G\cdot S_i)=B(G_i)=M\times S_i,$$

where $i=1$ or 2. Hence $X\epsilon M$. We find similarly that if $X\epsilon M$ then $X\epsilon B(G)$. We deduce that $M=B(G)$, a contradiction.

We may now suppose $M$ connected.

Suppose $dM=-1$. Then $|E|=0$ or 1 since $M$ is connected. We thus have $M=B(G)$ for some graph $G$ with just one vertex. If $|E|=0$, $G$ is edgeless. If $|E|=1$ $G$ has a single loop.

If $dM=0$ we have $M=\{E\}$, by 3.57 and the connection of $M$. Then $M=B(G)$, where $G$ has just two vertices and each cell of $M$ is an edge incident with each of them.

Next suppose $dM=1$. Then $E$ is a connected line of $M$. Let its three points be $X_1$, $X_2$, and $X_3$. Each is the mod 2 sum of the others, by 7.12. Hence they have no common cell. The sets $X_1\cap X_2$, $X_2\cap X_3$, and $X_3\cap X_1$ are non-null, by Axiom I, and their union is $E$. Let $G$ be a graph with just three vertices $v_1$, $v_2$, and $v_3$, and such that $E(G)=E$. We take each member of $X_i\cap X_j$ to have ends $v_i$ and $v_j$, $(1\leq i<j\leq 3)$. It is readily verified that $M=B(G)$.

We deduce that $dM\geq 2$.

Now $M$ has no minor of Type BI, BII, KI, or KII, by 7.51. Hence there is an atom $Y$ of $M$ having two distinct bridges, by 8.63. Moreover $M$ is even, by 9.31, and therefore $Y$ is bridge-separable. We can arrange the bridges of $Y$ in $M$ in two disjoint classes $P$ and $Q$ so that no two members of the same class overlap, and we can clearly arrange that $P$ and $Q$ are non-null.

Let $S$ and $T$ be the unions of the members of $P$ and $Q$ respectively. The matroids $M\times(S\cup Y)$ and $M\times(T\cup Y)$ are connected, by 3.48 and 8.13. For if the $S$ of 8.13 is a bridge of $Y$ the condition $(M\cdot(E-Y))\times S=M\times S$ is equivalent to $M\cdot S=M\times S$. (See 3.42.)

There exist graphs $H$ and $K$ such that

$$B(H)=M\times(S\cup Y),$$

$$B(K)=M\times(T\cup Y),$$

by the choice of $M$.

Hence we may choose $H$ to have a vertex $h$ which is incident with the members of $Y$ and no other edges, by 9.25 and 8.53. Similarly $K$ may be supposed to have a vertex $k$ incident only with the members of $Y$.

We may take $V(H)$ and $V(K)$ to be disjoint subsets of some larger set. We construct a graph $G$ as follows

$$E(G)=E$$
$$V(G)=(V(H)-\{h\})\cup(V(K)-\{k\}).$$

If $A\epsilon E-Y$ the ends of $A$ in $G$ are its ends in $H$ or $K$. But if $A\epsilon Y$ then its ends in $G$ are its end other than $h$ in $H$ and its end other than $k$ in $K$. We illustrate this construction in figure 9K.

There is a binary chain-group $N$ such that $M=M(N)$.

Let $f$ be any chain of $N$. Its restriction to $E-Y$ is a sum of elementary chains $f_1, f_2, \ldots, f_s$ of $N\cdot(E-Y)$, by 5.32. Each $f_i$ has a domain restricted to one of the bridges in $Y$ in $M$. Hence $f_i$ is the restriction to $E-Y$ of a chain $g_i$ of $N$ such that $||g_i||$ is contained in $S\cup Y$ or $T\cup Y$. The sum of $f$ and the chains $g_i$ has a domain contained in $Y$. It is thus either zero or the elementary chain of $N$ with domain $Y$.

We conclude that $f$ is the sum of two chains $p$ and $q$ of $N$ with domains contained in $S\cup Y$ and $T\cup Y$ respectively.

The restriction of $p$ to $S\cup Y$ is a coboundary of a 0-chain $l$ of $H$ over $GF(2)$. We may change each coefficient in $l$ by 1 without altering $\delta l$, by the definition of a coboundary. We may therefore suppose $l(h)=0$. But then $p$ is the coboundary of a corresponding 0-chain in $G$. Similarly $q$ is a coboundary in $G$. Hence $f$ is such a coboundary.

Conversely suppose $f$ is a coboundary of a 0-chain $l$ in $G$. Let $l_1$ and $l_2$ be formed from $l$ by replacing the coefficients of the vertices in $V(H)-\{h\}$ and $V(K)-\{k\}$ respectively by zeroes. Then $\delta l=\delta l_1+\delta l_2$. Moreover $\|\delta l_1\|\subseteq S\cup Y$ and $\|l_2\|\subseteq T\cup Y$. The restrictions to $S\cup Y$ and $T\cup Y$ of $\delta l_1$ and $\delta l_2$ are chains of $N\times(S\cup Y)$ and $N\times(T\cup Y)$ respectively. Hence $\delta l_1$ and $\delta l_2$ are chains of $N$, and therefore $\delta l$ is a chain of $N$.

We deduce that $N$ is the coboundary group of $G$. Hence $M=M(N)=B(G)$, contrary to the choice of $M$. The theorem follows.

We may now combine some of the preceding results into the following theorem.

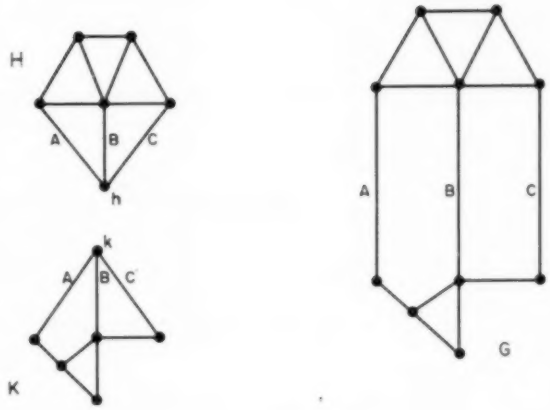



FIGURE 9K

9.42 *A binary matroid is graphic if and only if it is regular and has no minor of Type KI or KII.*

PROOF: If $M$ satisfies this condition it is graphic, by 9.41. Conversely if $M$ is graphic it is regular, by 2.54. No minor of $M$ is of Type KI or KII, by 5.62, for matroids of these two types are odd, by 8.31, and therefore not graphic, by 9.23.

Cographic matroids can be characterized by a dual form of 9.42. They are those regular matroids for which no minor is the bond-matroid of a Kuratowski graph.

A matroid is *planar* if it is both graphic and cographic. It is possible to derive Kuratowski's Theorem on planar graphs by applying 9.42 to the polygon-matroid of a planar graph.

An algorithm for determining whether a given binary matroid, given by a representative matrix, is graphic has recently been presented in the Canadian Journal of Mathematics **16,** 108–127 (1964).

(Paper 69B1–131)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Menger's Theorem for Matroids

W. T. Tutte

Professor of Mathematics, University of Waterloo, Ontario, Canada

(November 16, 1964)

Menger's Theorem asserts that if $x$ and $y$ are vertices of a graph which are not joined by an edge and if it takes at least $k$ other vertices to separate $x$ and $y$, then $x$ and $y$ can be joined by $k$ distinct arcs in the graph which have only their end-vertices in common.

Much of the paper is taken up by the determination of an equivalent statement involving the ranks of subgraphs of contractions of the original graph, but not explicitly mentioning vertices. An analogous proposition for matroids can then be stated. For matroid theory has analogs of subgraphs and contractions, but not of vertices. Once stated the generalized theorem can be proved without great difficulty.

## 1. Statement of the Theorem

In this paper we state and prove a theorem about general matroids which can be regarded as a generalization of Menger's Theorem for graphs. We state this theorem in the present section. The remainder of the paper falls into two parts. The first part, constituting section 2, is concerned with justifying the claim that Menger's Theorem can be deduced from our matroid theorem as a special case. The second part is concerned with the proof of the theorem. Thus a reader interested in general matroids but not in graphs might omit section 2.

Let $M$ be a matroid on a set $E$, and let $X$ and $Y$ be disjoint subsets of $E$.

If $\{S, T\}$ is any partition of $E$ we write

$$\xi(M, S, T) = r(M) - r(M \times S) - r(M \times T) + 1.$$

We also write $k(M; X, Y)$ for the minimum of $\xi(M; S, T)$, taken over all partitions $\{S, T\}$ of $E$ for which $X \subseteq S$ and $Y \subseteq T$.

We define $h(M; X, Y)$ as the maximum of $\xi(M'; X, Y)$, taken over all minors $M'$ of $M$ of the form $(M \times S) \cdot (X \cup Y)$.

The theorem to be proved is

$$h(M; X, Y) = k(M; X, Y). \tag{1.1}$$

## 2. Menger's Theorem for Graphs

Let $x$ and $y$ be distinct vertices of a graph $G$ which are not joined by an edge. A set $S \subseteq V(G) - \{x, y\}$ is said to *separate* $x$ and $y$ if each path from $x$ to $y$ in $G$ passes through a vertex of $S$. The least possible number $|S|$ of members of such a set $S$ will be denoted by $k(G; x, y)$.

A path in $G$ is called *simple* if it passes through no edge or vertex more than once.

Let $\{P_1, P_2, \ldots, P_h\}$ be a set of $h$ simple paths in $G$. They are *disjoint* if no two have a common edge or vertex. If they are all paths from $x$ to $y$ we say they are *internally disjoint* if no two of them have a common edge or vertex apart from the two vertices $x$ and $y$.

We write $h(G; x, y)$ for the maximum number of internally disjoint simple paths in $G$ from $x$ to $y$. Menger's Theorem, in one version, asserts that

$$h(G; x, y) = k(G; x, y). \tag{2.1}$$

The theorem can be stated in several equivalent though apparently different forms. The most convenient for our purposes is the following.

We replace $x$ and $y$ by disjoint non-null sets $X$ and $Y$ of edges of $G$. We define $k(G; X, Y)$ as the smallest number of vertices forming a set $S$ which separates $X$ and $Y$, that is such that each simple path in $G$ with one end a vertex of $G \cdot X$ and the other of $G \cdot Y$ passes through a vertex of $S$. (Here we use the notation of sec. 3.1 of "Lectures on Matroids".)

We now define $h(G; X, Y)$ as the maximum number of disjoint simple paths joining $X$ and $Y$, that is having one end in $V(G \cdot X)$ and the other in $V(G \cdot Y)$. It must be borne in mind that a "simple" path may be degenerate, that is consist of a single vertex.

Our second version of Menger's Theorem asserts that

$$h(G; X, Y) = k(G; X, Y). \tag{2.2}$$

To deduce 2.1 from 2.2 we take $X$ and $Y$ to be the sets of edges incident with $x$ and $y$ respectively. The derivation of 2.2 from 2.1 becomes obvious after the following construction is made. We introduce new vertices $x$ and $y$, and join them to all the vertices of $G \cdot X$ and $G \cdot Y$ respectively by new edges.

It might seem more natural to take $X$ and $Y$ as sets of vertices rather than edges. This is done, for example, by Kōnig. But then no generalization to matroids is suggested, since matroids have no vertices.

In discussions of Theorem 2.2 there is no real loss of generality in supposing $G \cdot X$ to be connected. For we can make it so by adjoining new edges, each to both $G$ and $X$, and this operation does not affect the numbers $h(G; X, Y)$ and $k(G; X, Y)$. There is also no harm in assuming $G \cdot Y$ connected.

We may also assume $G$ connected. For in all non-trivial cases the connected subgraphs $G \cdot X$ and $G \cdot Y$ are contained in the same component $H$ of $G$, and we may replace $G$ by $H$.

Let us try to give a definition of $k(G; X, Y)$ which does not mention vertices.

If $S$ and $T$ are complementary subsets of $E(G)$ we define $\eta(G; S, T)$ as the number of common vertices of $G \cdot S$ and $G \cdot T$. We next define $\mu(G; X, Y)$ as the minimum value of $\eta(G; S, T)$, when $S$ and $T$ are restricted only by the conditions $X \subseteq S$ and $Y \subseteq T$.

2.3 $$\mu(G; X, Y) = k(G; X, Y).$$

PROOF: If $S$ and $T$ are defined as above, then each simple path joining $X$ and $Y$ has its first vertex incident with a member of $S$ and its last with a member of $T$. Hence one of its vertices is incident with members of both $S$ and $T$. We deduce that the common vertices of $G \cdot S$ and $G \cdot T$ form a set separating $X$ and $Y$. Thus $\mu(G; X, Y) \geqslant k(G; X, Y)$.

Conversely suppose we have a set $Q$ of vertices separating $X$ and $Y$. Let $S$ be the set of all edges $A$ of $G$ such that some simple path $P$ traversing $A$ has the following properties.

(i) The first vertex of $P$ is incident with an edge of $X$.

(ii) No vertex of $P$, save possibly the last, belongs to $Q$.

It is clear that $S \cap Y$ is null and that if $T = E(G) - S$ then any common vertex of $G \cdot S$ and $G \cdot T$ is in $Q$. Hence $\mu(G; X, Y) \leqslant k(G; X, Y)$, since as a special case we may have $|Q| = k(G; X, Y)$.

The theorem follows.

2.4 *Suppose $G$, $G \cdot X$ and $G \cdot Y$ are connected. Let $\{S, T\}$ be a partition of $E(G)$ such that $X \subseteq S$ and $Y \subseteq T$. Then*

$$\eta(G; S, T) - k(G; X, Y) \geqslant p_0(G \cdot S) + p_0(G \cdot T) - 2.$$

PROOF: Assume that the theorem fails, and that $p_0(G \cdot S) + p_0(G \cdot T)$ has the least value consistent with this condition.

Clearly $p_0(G \cdot S) + p_0(G \cdot T) \geqslant 3$, by 2.3.

Without loss of generality we may suppose $G \cdot S$ to have a component $G \cdot S_1$ which does not meet $G \cdot X$. Let this component meet $q$ components of $G \cdot T$ and let it have $p$ common vertices with $G \cdot T$. Then, abbreviating $p_0(G \cdot U)$ as $p_0(U)$, we have the following relationships.

(i) $0 < q \leqslant p$.

(ii) $p_0(S - S_1) = p_0(S) - 1$

(iii) $p_0(T \cup S_1) = p_0(T) - q + 1$,

(iv) $\eta(G; S - S_1, T \cup S_1) = \eta(G; S, T) - p$.

The inequality $0 < q$ follows from the fact that $G$ is connected.

We now have

$$\begin{aligned} &\eta(G; S, T) - k(G; X, Y) \\ &= \eta(G; S - S_1, T \cup S_1) - k(G; X, Y) + p, \\ &\geqslant p_0(S - S_1) + p_0(T \cup S_1) + p - 2, \end{aligned}$$

by the choice of $S$ and $T$. For

$$p_0(S - S_1) + p_0(T \cup S_1) < p_0(S) + p_0(T),$$

by (i), (ii), and (iii). Hence

$$\begin{aligned} &\eta(G; S, T) - k(G; X, Y) \\ &\geqslant p_0(S) + p_0(T) + p - q - 2, \end{aligned}$$

by (ii) and (iii),

$$\geqslant p_0(S) + p_0(T) - 2.$$

But this is contrary to assumption. The theorem follows.

If $G$ is a graph of $p_0(G)$ components the rank of its bond-matroid $B(G)$ can be deduced from Theorem 2.21 of "Matroids." A principal forest of $G$ consists of $p_0(G)$ disjoint trees, and in any tree the number of vertices exceeds the number of edges by 1. So by the theorem just quoted we have

2.5 $$r(B(G)) = |V(G)| - p_0(G).$$

Now let $S$ and $T$ be complementary subsets of $E(G)$. Using Theorem 3.321 of "Matroids" we find that

2.6 $$\begin{aligned} &r(B(G)) - r(B(G) \times S) - r(B(G) \times T) \\ &= r(B(G) \cdot T) + r(B(G) \cdot S) - r(B(G)), \end{aligned}$$

by "Matroids," 3.54,

$$= V(G \cdot T) + V(G \cdot S) - V(G)$$

$$- p_0(G \cdot T) - p_0(G \cdot S) + p_0(G),$$

by "Matroids," 3.321,

$$= \eta(G; S, T) + p_0(G)$$

$$- p_0(G \cdot S) - p_0(G \cdot T),$$

if $G$ has no isolated vertices.

2.7 *If $G$, $G \cdot X$ and $G \cdot Y$ are connected, then $k(G; X, Y)$ is the minimum of*

$$r(B(G)) - r(B(G) \times S) - r(B(G) \times T) + 1,$$

*taken over all partitions $\{S, T\}$ of $E(G)$ such that $X \subseteq S$ and $Y \subseteq T$.*

PROOF: Denote this minimum by $m$. Applying 2.4 to the result just proved we find that $m \geq k(G; X, Y)$. Moreover if $\eta(G; S, T) = k(G; X, Y)$ we have $p_0(G \cdot S) = p_0(G \cdot T) = 1$, by 2.4, and therefore

$$r(B(G)) - r(B(G) \times S) - r(B(G) \times T) + 1 = \eta(G; S, T) = k(G; X, Y).$$

The theorem follows.

Theorem 2.7 has the following corollary

2.71 $$k(G; X, Y) = k(B(G); X, Y).$$

It is clear that the definition of $h(G; X, Y)$ given above is equivalent to the following; $h(G; X, Y)$ is the maximum number of disjoint connected subgraphs of $G$, no one including an edge of $X \cup Y$, such that each has at least one vertex in common with each of the graphs $G \cdot X$ and $G \cdot Y$.

We express this result in terms of the notation of section 3.1 of "Matroids" as follows. The vertices of the graph $H = (G \times S) \cdot (X \cup Y)$, where $X \cup Y \subseteq S \subseteq E$, are the components of $G : (E(G) - S)$. They are thus disjoint connected subgraphs of $G$. The edges of $H$ are the members of $X \cup Y$. The ends of each such edge in $H$ are those vertices of that graph which, as subgraphs of $G$, contain the ends of the edge in $G$.

It follows that $h(G; X, Y)$ is the maximum number of common vertices of $H \cdot X$ and $H \cdot Y$, where $H = (G \times S) \cdot (X \cup Y)$ ranges over all possible choices of $S$. It is the maximum of $\eta(H; X, Y)$, taken over all possible choices of $S$. If we assume that $G$, $G \cdot X$ and $G \cdot Y$ are connected we can readily verify that $H$, $H \cdot X$ and $H \cdot Y$ are also connected. We then have $\eta(H; X, Y) = \xi(B(H); X, Y)$, by 2.6. But $B(H) = (B(G) \times S) \cdot (X \cup Y)$. It follows that

2.8 $$h(B(G); X, Y) = h(G; X, Y).$$

Formulae 2.71 and 2.8 hold when $G$, $G \cdot X$ and $G \cdot Y$ are connected. They justify the assertion that 1.1 then includes 2.2 as a special case. The extension to the general case for graphs may be made by devices already described.

## 3. Some Auxiliary Theorems

We prove Menger's Theorem for matroids in the next section. Here we give a preliminary discussion of the functions

$$\xi(M; S, T) \text{ and } k(M; X, Y).$$

3.1 *Let $\{S, T\}$ be any partition of $E$. Then*

$$\xi(M^*; S, T) = \xi(M; S, T).$$

PROOF:

$$\xi(M; S, T)$$

$$= r(M) - r(M \times S) - r(M \times T) + 1$$

$$= - r(M) + r(M \cdot T) + r(M \cdot S) + 1,$$

by "Matroids," 3.54,

$$= - |E| + r(M^*) + |T| - r((M \cdot T)^*) + |S| - r((M \cdot S)^*) + 1,$$

by "Matroids," 3.52,

$$= r(M^*) - r(M^* \times T) - r(M^* \times S) + 1,$$

by "Matroids," 3.351,

$$= \xi(M^*, S, T).$$

As a Corollary we have

3.11 *Let $X$ and $Y$ be disjoint subsets of $E$. Then*

$$k(M^*; X, Y) = k(M; X, Y).$$

3.2 *Let $\{S, T\}$ and $\{U, V\}$ be partitions of $E$. Then*

$$\xi(M; S \cup U, T \cap V) + \xi(M; S \cap U, T \cup V) \leq \xi(M; S, T) + \xi(M; U, V).$$

PROOF:

$$\xi(M: S \cup U, T \cap V) + \xi(M; S \cap U, T \cup V)$$

$$= 2r(M) + 2 - r(M \times (S \cup U)) - r(M \times (S \cap U))$$

$$- r(M \times (T \cup V)) - r(M \times (T \cap V))$$

$$\leq 2r(M) + 2 - r(M \times S) - r(M \times U)$$

$$- r(M \times T) - r(M \times V),$$

by "Matroids," 3.56

$$= \xi(M;\, S,\, T) + \xi(M;\, U,\, V).$$

Suppose $a \epsilon S \subseteq E$. We define an integer $\theta(M;\, S,\, a)$ as follows. $\theta(M;\, S,\, a) = 1$ if there is an atom $X$ of $M$ such that $a \epsilon X \subseteq S$, and $\theta(M;\, S,\, a) = 0$ otherwise. Thus

$$\theta(M;\, S,\, a) = r((M \times S) \cdot \{a\}).$$

We take note of the following trivial result.

3.3 *If* $a \epsilon S \subseteq T \subseteq E$, *and* $\theta(M;\, S,\, a) = 1$, *then* $\theta(M;\, T,\, a) = 1$.

From now on we abbreviate $r(M \times S)$ as $r(S)$. Thus, using 3.54 of "Matroids" we have

3.4 *If* $a \epsilon S \subseteq E$, *then*

$$r(S) = r(S - \{a\}) + \theta(M;\, S,\, a)$$

Given a cell $a \epsilon E$ we abbreviate the expressions $M \times (E - \{a\})$ and $M \cdot (E - \{a\})$ as $M'_a$ and $M''_a$ respectively.

3.51 *If* $\{S,\, T\}$ *is a partition of* $E - \{a\}$ *then*

$$\xi(M'_a;\, S,\, T)$$

$$= \xi(M;\, S \cup \{a\},\, T) - \theta(M;\, E,\, a) + \theta(M,\, S \cup \{a\},\, a),$$

by 3.4 and the definition of the function $\xi(M;\, U,\, V)$.

3.52 *If* $\{S,\, T\}$ *and* $a$ *are as in* 3.51, *then*

$$\xi(M''_a;\, S,\, T)$$

$$= \xi(M;\, S \cup \{a\},\, T) + r\{a\}) - \theta(M;\, T \cup \{a\},\, a).$$

PROOF:

If $U \subseteq E - \{a\}$ we have

$$r(M''_a \times U) = r((M \times (U \cup \{a\})) \cdot U),$$

by "Matroids," 3.333,

$$= r(U \cup \{a\}) - r(\{a\}),$$

by "Matroids," 3.54.

We deduce that

$$\xi(M''_a;\, S,\, T)$$

$$= r(M''_a) - r(M''_a \times S) - r(M''_a \times T) + 1$$

$$= r(E) - r(\{a\}) - r(S \cup \{a\}) + r(\{a\})$$

$$- r(T \cup \{a\}) + r(\{a\}) + 1$$

$$= \xi(M;\, S \cup \{a\},\, T) + r(\{a\})$$

$$- \theta(M;\, T \cup \{a\},\, a).$$

By 3.3 the difference $\theta(M;\, E,\, a) - \theta(M;\, S \cup \{a\},\, a)$ of 3.51 is either 1 or 0. Similarly, since $r(\{a\}) = \theta(M;\, \{a\},\, a)$, the difference $\theta(M;\, T \cup \{a\},\, a) - r(\{a\})$ of 3.52 is either 1 or 0. Hence, by 3.51, 3.52, and the definition of $k(M;\, X,\, Y)$ we have the following theorem.

3.6 *If* $X$ *and* $Y$ *are disjoint subsets of* $E$, *and* $a \epsilon E - (X \cup Y)$, *then the following propositions hold.*

(i) $k(M;\, X,\, Y) \geqslant k(M'_a;\, X,\, Y) \geqslant k(M;\, X, Y) - 1$,

(ii) $k(M;\, X,\, Y) \geqslant k(M'_a;\, X,\, Y) \geqslant k(M;\, X,\, Y) - 1$.

We can improve upon this result as follows.

3.7 *If* $X$, $Y$ *and* $a$ *are as in* 3.6, *then either*

$$k(M'_a;\, X, Y) = k(M;\, X,\, Y)$$

or

$$k(M''_a;\, X,\, Y) = k(M;\, X,\, Y).$$

PROOF: Suppose the theorem to fail. Then there exist partitions $\{S,\, T\}$ and $\{U,\, V\}$ of $E - \{a\}$, with $X \subseteq S \cap U$ and $Y \subseteq T \cap V$, which have the following properties.

(i) $\xi(M;\, S \cup \{a\},\, T) = \xi(M;\, S,\, T \cup \{a\})$

$$= \xi(M'_a;\, S,\, T) + 1 = k(M;\, X,\, Y),$$

(ii) $\xi(M;\, U \cup \{a\},\, V) = \xi(M;\, U,\, V \cup \{a\})$

$$= \xi(M''_a;\, U,\, V) + 1 = k(M;\, X,\, Y).$$

Using 3.51 we find

$$\xi(M'_a;\, U,\, V) \leqslant k(M;\, X,\, Y).$$

So, by (i) and 3.2,

$$\xi(M'_a;\, S \cup U,\, T \cap V) + \xi(M'_a;\, S \cap U,\, T \cup V)$$

$$\leqslant 2k(M;\, X,\, Y) - 1.$$

We may therefore suppose, without loss of generality, that

(iii) $\xi(M;\, S \cup U \cup \{a\},\, T \cap V)$
$= \xi(M;\, S \cup U,\, (T \cap V) \cup \{a\})$

$$= \xi(M'_a;\, S \cup U,\, T \cap V) + 1 = k(M;\, X,\, Y).$$

By 3.52 and (ii) we have

$$\theta(M;\, U \cup \{a\},\, a) = 1,$$

and by 3.51 and (iii) we have

$$\theta(M;\, S \cup U \cup \{a\},\, a) = 0.$$

But these results contradict 3.3. The theorem follows.

3.8 *If* $X$, $Y$, *and* $a$ *are as in* 3.6, *then*

$$h(M;\, X,\, Y) = \text{Max}\, [h(M'_a;\, X,\, Y),\, h(M'_a;\, X,\, Y)].$$

PROOF. Consider a minor $M_1=(M\times S)\cdot(X\cup Y)$ of $M$. If $a\notin S$ we can write $M_1=(M'_a\times S)\cdot(X\cup Y)$. If however $a\in S$ we have

$$M_1=((M\times S)\cdot(S-\{a\}))\cdot(X\cup Y)$$

$$=(M''_a\times(S-\{a\}))\cdot(X\cup Y).$$

Conversely a minor $M_2$ of $M'_a$ of the form $(M'_a\times S)\cdot(X\cup Y)$ can be rewritten as $(M\times S)\cdot(X\cup Y)$. For a minor $M_2=(M''_a\times S)\cdot(X\cup Y)$ we have

$$M_2=((M\cdot(E-\{a\}))\times S)\cdot(X\cup Y)$$

$$=(M\times(S\cup\{a\}))\cdot(X\cup Y).$$

Hence the minors of $M$ of the form $(M\times S)\cdot(X\cup Y)$ are the minors of $M'_a$ of the form $(M'_a\times S)\cdot(X\cup Y)$ together with those of $M''_a$ of the form $(M''_a\times S)\cdot(X\cup Y)$. The theorem follows.

(Paper 69B1–132)

## 4. Proof of the Main Theorem

We proceed to prove the proposition already stated as 2.1, namely that

4.1 $$h(M;X,Y)=k(M;X,Y)$$

for arbitrary disjoint subsets $X$ and $Y$ of $E$.

Write $m=|E-(X\cup Y)|$. If $m=0$ the theorem is trivially true, since both $h(M;X,Y)$ and $k(M;X,Y)$ are equal to $\xi(M;X,Y)$.

Assume as an inductive hypothesis that it is true whenever $m$ is less than some positive integer $q$ and consider the case $m=q$.

Choose $a\in E-(X\cup Y)$. Then

$$k(M;X,Y)=\text{Max}\,[k(M'_a;X,Y),\,k(M''_a;X,Y)],$$

by 3.6 and 3.7,

$$=\text{Max}\,[h(M'_a;X,Y),\,h(M''_a;X,Y)],$$

by the inductive hypothesis,

$$=h(M;X,Y),$$

by 3.8.

Thus the theorem holds also when $m=q$. It follows in general by induction.

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Single-Element Extensions of Matroids

**Henry H. Crapo**

**Assistant Professor of Mathematics, Northeastern University, Boston, Mass.**

(November 16, 1964)

Extensions of matroids to sets containing one additional element are characterized in terms of modular cuts of the lattice of closed subsets. An equivalent characterization is given in terms of linear subclasses of the set of circuits or bonds of the matroid. A scheme for the construction of finite geometric lattices is derived and the existence of at least $2^n$ nonisomorphic matroids on an $n$-element set is established.

## Contents

| | Page |
|---|---|
| 1. Introduction | 55 |
| 2. Differentials | 55 |
| 3. Unit increase functions | 57 |
| 4. Exact differentials | 58 |
| 5. Single-element extensions | 59 |
| 6. Linear subclasses | 60 |
| 7. Geometric lattices | 62 |
| 8. Numerical classification of matroids | 62 |
| 9. Bibliography | 64 |

## 1. Introduction [1]

In order to facilitate inductive proofs of matroid[2] theorems, we shall set forth a characterization of single-element extensions of matroid structures. Our techniques are those developed in our recent paper, "Lattice Differentials and the Theory of Combinatorial Independence". Sufficient material is included in the next three sections to support the proofs of the extension theorems. The reader is referred to the above paper for details.

A matroid, defined on a finite set, may be thought of as a structure defined on a lattice, i.e., on the Boolean algebra of all subsets of the set. Not all the properties of Boolean algebras are used in the proof of matroid theorems, however. We have chosen modular lattices as the proper domain for our presentation of the material needed for extension theory.

## 2. Differentials

An element $y$ is said to *cover* an element $x$ in a lattice $L$ if $x < y$, and for any element $z$, $x < z \leq y$, implies $z = y$. A lattice $L$ is *modular* if, for all elements $x$ and $y$ in $L$, $x$ covers $x \wedge y$ if and only if $x \vee y$ covers $y$. A sequence $p$: $x = p_0 < p_1 < \ldots < p_n = y$ of elements $p_i$ of a lattice $L$, in which $p_i$ covers $p_{i-1}$ for $i = 1, \ldots, n$, is a *path* (of length $n$) from $x$ to $y$. A *step* is a path of length 1.

The partial order of the lattice induces an order on the set of steps of the lattice. If $[x, y]$ and $[u, w]$ are steps of a lattice $L$, we say $[x, y] \leq [u, w]$ if and only if $x = y \wedge u$ and $y \vee u = w$. If, in a modular lattice, a step $[u, w]$ covers a step $[x, y]$ in this ordering, then $[u, x]$ and $[y, w]$ are also steps.

Definition: A function $R$ defined for all steps $[x, y]$ of a finite modular lattice $L$, and taking values in the two-element lattice[3] $\{0, 1\}$, is a *differential* if and only if it is

(a) Projective on steps: If steps $[x, y]$ and $[u, w]$ are in the order $[x, y] \leq [u, w]$, then $R[x, y] \geq R[u, w]$.

(b) Subadditive on steps: If a step $[u, w]$ covers a step $[x, y]$ then $R[x, y] \leq R[x, u] \vee R[u, w]$.

The local character of differentials may be represented pictorially. A *local graph* is any assignment of values 0 or 1 to the four steps of an interval $[x \wedge y, x \vee y]$ in a finite modular lattice $L$, where the elements $x$ and $y$ cover $x \wedge y$ and are thus covered by $x \vee y$.

PROPOSITION: *An assignment of values* 0 *and* 1 *to the steps of a finite modular lattice* L *determines a differential if and only if all local graphs are of the following five types: zero, mixed, prime, one, or inexact.* (These types of local graph are defined as figure 1. We indicate steps on which $R = 1$ by double lines, $R = 0$ by single lines.)

[1] The statement of the extension theorem and the classification of matroids are taken from the author's Ph.D. thesis, entitled "On the Theory of Combinatorial Independence", written under the supervision of Professor Gian-Carlo Rota, and presented to the Mass. Inst. of Technology, May 1964. We are grateful to Jack Edmonds, National Bureau of Standards, who pointed out the significance of these theorems and encouraged us in their presentation.

[2] H. Whitney, Amer. J., 1935.

[3] Differentials of higher order may be obtained by allowing the range of the differential to be a lattice more general than $\{0, 1\}$.

FIGURE 1

PROOF: The projective and subadditive properties of differentials eliminate all local graphs except the five enumerated above. Conversely, if $R$ is a function on steps with values in the lattice $\{0, 1\}$, and if all local graphs of $R$ are zero, mixed, prime, one, or inexact, the subadditive property follows, and the projective property is established, for steps $[x, y]$ and $[u, w]$, with $[x, y] \leq [u, w]$, as follows. Let $x = p_0 < p_1 < \ldots < p_n = u$ be a path from $x$ to $u$, and let $q = p \vee y$ be the projected path $y = p_0 \vee y < p_1 \vee y < \ldots < p_n \vee y = w$ from $y$ to $w$. If $R[x, y] < R[u, w]$, then $R[x, y] = 0$ and $R[u, w] = 1$. Let $p_i$ be the greatest element such that $R[p_i, q_i] = 0$. Then $R[p_{i+1}, q_{i+1}] = 1$. Such a local graph $p_i, q_i, p_{i+1}, q_{i+1}$ is neither zero, mixed, prime, nor inexact.

A differential $R$ is termed *exact* if and only if it has no inexact local graph. Exact differentials are the subject of section 4, below.

The three principal sources of differentials are closure operators, join-homomorphisms, and finite lattices. We describe these in turn.

If $Cl$ is a closure operator ($x \leq Cl(x)$, and $x \leq Cl(y)$ implies $Cl(x) \leq Cl(y)$) on a finite modular lattice $L$,

$$R[x, y] = 1 \text{ if } Cl(x) < Cl(y)$$
$$= 0 \text{ otherwise}$$

defines a differential $R$ on the steps of the lattice $L$.

A differential is canonically related to a unique closure on its domain lattice. We say an element $x$ of a finite modular lattice $L$ is *closed* relative to a differential $R$ on $L$ if and only if $R[x, y] = 1$ for all steps of the form $[x, y]$. We shall prove closed elements are elements $x$ such that $Cl(x) = x$, in the closure $Cl$ associated with $R$.

PROPOSITION: *Existence of Closed Elements: If* R *is a differential on a finite modular lattice* L, *and if the differential* R *has value 1 on some step of* L, *there exists a closed element* z *in* L *other than the element 1.*

PROOF: Assume $R$ has value 1 on the step $[u, w]$, and say the lattice $L$ has height $n$. Choose a path $p$ from 0 to 1 which is maximal with respect to the number $k$ of initial steps on which the differential $R$ has value 0. Then all steps beginning at $p_k$ must have differential value 1, and $p_k$ is closed. It remains to prove that $k < n$, so that $p_k$ is not the uppermost element 1 of the lattice.

We prove by induction on the height $n$ of the lattice that a step $[u, w]$ on which $R$ has value 1 cannot coexist in $L$ with a path $p$ from 0 to 1 on which $R$ has value 0. If $n = 1$: $u = 0$ and $w = 1$, and no alternate path exists. Assume the truth of the statement for $n-1$, and let $L$ have height $n$. Since $L$ is modular, either $w \vee p_{n-1} = u \vee p_{n-1}$ and $w \wedge p_{n-1}$ covers $u \wedge p_{n-1}$, or else $w \vee p_{n-1}$ covers $u \vee p_{n-1}$ and $w \wedge p_{n-1} = u \wedge p_{n-1}$. In the first instance, we have the ordering of steps $[u \wedge p_{n-1}, w \wedge p_{n-1}] \leq [u, w]$, so $R[u \wedge p_{n-1}, w \wedge p_{n-1}] = 1$ in the lattice interval $[0, p_{n-1}]$ of height $n-1$. This step cannot coexist with the portion of the path $p$ from 0 to $p_{n-1}$. In the second instance, choose a path $q$ from $u$ to $p_{n-1}$, and let $j$ be the greatest index for which $R[q_j, w \vee q_j] = 1$. By subadditivity of $R$ on steps, $R[q_j, q_{j+1}] = 1$, where $[q_j, q_{j+1}]$ is also a step in the interval sublattice $[0, p_{n-1}]$ of height $n-1$.

PROPOSITION: *Closures Formed from Differentials: Given a differential* R *on a finite modular lattice* L,

$$\mathrm{Cl}(x) = \sup \{y;\ \exists \textit{ path from } x \textit{ to } y \textit{ along which } R = 0\}$$
$$= \inf \{z;\ z \textit{ closed}, x \leq z\}$$

*is a closure operator on the lattice* L, *the differential of which is* R.

PROOF: If an element $x$ is connected to two elements $y_1$ and $y_2$ by paths along which $R = 0$, every step in a path from $y_1$ to $y_1 \vee y_2$ exceeds, in the ordering on steps, some step in the path from $x$ to $y_2$. Thus $R = 0$ on the extended path from $x$ to $y_1 \vee y_2$, so $y_1 \vee y_2 \leq Cl(x)$. If a path exists from $x$ to $y$ along which $R = 0$, then $R$ must have value 0 on all steps of the lattice interval $[x, y]$, as was shown in the proof of the previous proposition. Thus the closure $Cl(x)$ may be characterized as the maximum element $y$ such that $R = 0$ everywhere on the lattice interval $[x, y]$. This element $Cl(x)$ must be closed, for otherwise a path along which $R = 0$ would exist from $x$ to an element covering $Cl(x)$.

The set of closed elements of the lattice $L$ is closed with respect to the lattice operation *inf*. If $x$ and $y$ are closed elements, and an element $u$ covers $x \wedge y$, then $u \not\leq x$ or $u \not\leq y$. Say $u \not\leq y$. Then $u \vee y$ covers $y$, $[x \wedge y, u] \leq [y, u \vee y]$ in the ordering on steps, and $R[x \wedge y, u] = 1$ because the element $y$ is closed and $R[y, u \vee y] = 1$. Thus, for any element $x$ in $L$, $Cl(x)$ may be characterized as the infimum of all closed elements $z$ such that $x \leq z$. As is proven in the theory of closure operators,[4] any *inf*-closed class of lattice elements generates a closure operator by this rule.

It remains to prove that the differential of the closure operator $Cl$ is the original differential $R$, i.e., that $R[x, y] = 1$ if and only if $Cl(x) < Cl(y)$. If $R[x, y] = 1$, there is a relatively closed element $z$ in the interval sublattice $[x, Cl(y)]$ other than $Cl(y)$. If an element $u$ covers $z$ in the lattice $L$, but $u \not\leq Cl(y)$, then $u \vee Cl(y)$ covers $Cl(y)$, $R[Cl(y), u \vee Cl(y)] = 1$, and $R[z, u] = 1$ because $R$ is projective on steps. Thus the element $z$ is closed, and $Cl(x) \leq z < Cl(y)$. Conversely, if $Cl(x) < Cl(y)$, then $y \vee Cl(x)$ covers $Cl(x)$ and $[x, y] \leq [Cl(x), y \vee Cl(x)]$ implies $R[x, y] = 1$.

The set of closed elements of a differential $R$ on a finite modular lattice $L$, with partial ordering induced by that on $L$, forms a finite lattice $L/R$, called the

[4] O. Ore, Theory of Graphs.

*lattice of closed elements of R on L.* The *meet-irreducible* elements (elements covered by exactly one other element) and *join-irreducible* elements (elements covering exactly one other element) figure prominently in the theory of differentials. Those meet-irreducible elements covered by 1 are called *coatoms;* those join-irreducible elements covering 0 are called *atoms.*

As a second source for differentials, consider join-homomorphisms $(f(x \vee y) = f(x) \vee f(y))$ from a finite modular lattice $L$ into a lattice $Q$. A closure operator and a differential are induced on the domain of a join-homomorphism in much the same way that a partition is induced on the domain of a function.

PROPOSITION: *If f is a join-homomorphism from a finite modular lattice* L *into a lattice* Q, *then*

$$R[x, y] = 1 \text{ if } f(x) < f(y)$$

$$R[x, y] = 0 \text{ otherwise}$$

*defines a differential* R *on the steps* [x, y] *of the lattice* L.

PROOF: Assume steps $[x, y]$ and $[u, w]$ are in the order $[x, y] \leq [u, w]$. If $R[x, y] = 0$, then $f(y) = f(x) \leq f(u)$, because $x \leq u$ and $f$, being a join-homomorphism, is order-preserving. Thus $f(w) = f(y \vee u) = f(y) \vee f(u) = f(u)$, so $R[u, w] = 0$, proving $R$ is projective on steps. If a step $[u, w]$ covers a step $[x, y]$, and if $R[x, u] = R[u, w] = 0$, then $f(x) \leq f(y) \leq f(w) = f(u) = f(x)$, so $f(x) = f(y)$, and $R[x, y] = 0$, proving subadditivity.

Any differential may be represented as that of a join-homomorphism, by letting the lattice $Q$ be the Boolean algebra of all subsets of the set $M$ of meet-irreducible elements of the lattice $L/R$, and by mapping each element $x \epsilon L$ into the subset of $M$ composed of meet-irreducible elements of $L/R$ not above $Cl(x)$. The proof is available in "Lattice Differentials . . .".

A third source of differentials is the theory of finite lattices. Let $Q$ be an arbitrary finite lattice, let $A$ be the set of join-irreducible elements of $Q$ (those elements covering exactly one other element), and let $L$ be the Boolean algebra of all subsets of the set $A$. Define a closure operator $Cl$ on the Boolean algebra $L$ by mapping every subset $x$ of the set $A$ into the subset $Cl(x)$ containing all join-irreducible elements $e$ of $Q$ such that $e \leq sup\ d$, $d \epsilon x$ the supremum being taken in the lattice $Q$. The differential of this closure operator is the *structure differential* of the lattice $Q$.

These three examples indicate the scope of the theory of differentials, a theory coextensive with the theories of finite lattices and of closure operators on finite modular lattices.

## 3. Unit Increase Functions

Before beginning a discussion of exact differentials, let us consider systems satisfying Whitney's first rank function axiom for matroids. Called unit increase functions, they provide a shortcut to exact differentials, a shortcut used in the single-element extension theorem. Differences of unit increase functions, like differentials, may be characterized by conditions on local graphs.

An integer-valued function $r$ defined on a finite modular lattice $L$ is a *unit increase function* if $r(y) - r(x)$ is either 0 or 1 whenever an element $y$ covers an element $x$ in $L$. The function $R$ defined for all steps $[x, y]$ in $L$ by

$$R[x, y] = 1 \text{ if } r(x) < r(y)$$

$$= 0 \text{ otherwise}$$

is the *difference* of the unit increase function $r$. Notice that for all steps $[x, y]$, $R[x, y] = r(y) - r(x)$. In drawing comparisons with differentials, we think of the values 0 and 1 of $R$ as elements of the lattice $\{0, 1\}$. The local characterization of differences of unit increase functions is as follows.

PROPOSITION: *Let* L *be a finite modular lattice. An assignment of values 0 and 1 to the steps of the lattice* L *is that of a difference of a unit increase function on* L *if and only if all local graphs are of the following five types: zero, mixed, prime, one or inverted.* (See fig. 2).

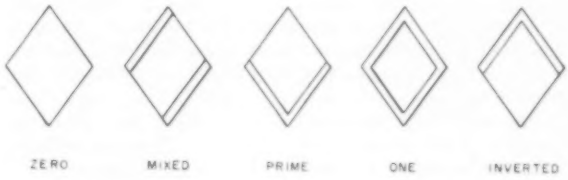



FIGURE 2

PROOF: It is clear that the five possible local graphs not listed above may not appear in a difference of a unit increase function. Given an assignment $R$ of values 0 and 1 to the steps of $L$ such that all local graphs are zero, mixed, prime, one or inverted we construct a unit increase function $r$ by setting

$$r(x) = \sum_{i=1}^{n} R[p_{i-1}, p_i], \text{ where } p \text{ is any path from 0 to } x.$$

That $r(x)$ is independent of path is established by an inductive argument. Say $x$ is of height $n$ in the lattice $L$, and all paths of length $n-1$ or less have the same $R$-sum. Let $p$: $0 = p_0 < p_1 < \ldots < p_n = x$ and $q$: $0 = q_0 < q_1 < \ldots < q_n = x$ be two paths from 0 to $x$. If $p_1 = q_1$, the $R$-sums are equal on $p$ and $q$. If $p_1 \neq q_1$, $0 = p_1 \wedge q_1$ and $p_1 \vee q_1$ covers $p_1$ and $q_1$. Form a path $s$ from $p_1 \vee q_1$ to $x$. The $R$-sum along $p$ is equal to that to $p_1$ and $p_1 \vee q_1$, thence along $s$, by the induction assumption. This sum is in turn equal to the sum to $q_1$ and $p_1 \vee q_1$, thence along $s$, because the local graph on 0, $p_1$, $q_1$, $p_1 \vee q_1$ is one of the five types listed above. But this sum is equal to that along the path $q$, by the induction assumption. The $R$-sum, thus well-defined, is a unit increase function, because $r(y) - r(x) = R[x, y] = 0$ or 1, for all steps $[x, y]$.

This completes the local characterization of differences of unit increase functions on finite modular lattices. A unit increase function whose difference has no inverted local graph is termed a *Whitney rank function.*

We find the following example due to Edmonds of some importance. Given a finite set, and a class of distinguished subsets called *independent* sets, a class closed with respect to taking of subsets, we may define the *rank* of a subset to be the number of elements in its largest independent subset. This rank is a unit increase function. The single assumption made, that every subset of an independent set is independent, is the assumption defining simplicial complexes in algebraic topology.

## 4. Exact Differentials

An exact differential was defined as a differential, all of whose local graphs are zero, mixed, prime, or one.

PROPOSITION: *Exact Differentials: If* r *is a unit increase function on a finite modular lattice and if the difference* R *of* r *is a differential, then* r *is a Whitney rank function, and* R *is an exact differential.*

PROOF: If $r$ is a unit increase function with difference $R$, then all local graphs of $R$ are zero, mixed, prime, one, or inverted. If $R$ is also a differential, all local graphs are zero, mixed, prime, one, or inexact. Thus all local graphs are zero, mixed, prime, or one, $r$ is a Whitney rank function, and $R$ is an exact differential.

PROPOSITION: *Matroids: Let* L *be the Boolean algebra of all subsets of a finite set. Then an integer-valued function* r *for which* $r(\Phi)=0$ *is a Whitney rank function if and only if* r *is the rank function of a matroid.*

PROOF: Whitney's first axiom for matroids in terms of rank functions was taken as the definition of a unit increase function. If $r(x)=r(x\vee e)=r(x\vee f)$ for elements $e$ and $f$ not in a subset $x$, then $x$, $x\vee e$, $x\vee f$, $x\vee e\vee f$ form a local graph, on which $R[x, x\vee e]=R[x, x\vee f]=0$. If $R$ is exact, this local graph must be zero, so $R[x\vee e, x\vee e\vee f]=R[x\vee f, x\vee e\vee f]=0$, and thus $r(x)=r(x\vee e\vee f)$, which gives Whitney's second axiom for matroids in terms of rank functions. Conversely, if $r$ is the rank function of a matroid, no local graph of its difference $R$ may be inverted, so $r$ is a Whitney rank function.

Several matroid theorems apply without alteration to exact differentials. For example, the statement "if an element $e$ is dependent upon a subset $x$ but upon no smaller subset, then $x$ is a circuit" may be phrased in terms of closed elements, as follows.

PROPOSITION: *Characterization of Meet-Irreducible Closed Elements: Given a step* [x, y] *in a modular lattice* L *and an exact differential* R *defined on* L, *the element* x *is meet-irreducible in the lattice* L/R *of closed elements if and only if the step* [x, y] *is maximal, in the ordering on steps, among steps on which the differential* R *has value 1. (The proof may be found in "Lattice Differentials . . .".)*

The concept of matroid duality is available for exact differentials. An element $x$ in the domain lattice $L$ of a differential $R$ is *dual-closed* if $R[y, x]=0$ for all steps of the form $[y, x]$. The class of dual-closed elements is closed under the lattice operation *sup;* the image of this class in the inverted lattice $L$ is closed under the lattice operation *inf*, and gives rise to a closure operator and a differential $R^*$ on the inverted lattice $L$. It may be shown that $R^{**}=R$ if $R$ is an exact differential.

Several conditions on differentials, equivalent to the exactness condition, will support the proof of the extension theorem. These conditions are given below; their equivalence is proven in "Lattice Differentials . . .".

PROPOSITION: *Equivalent Exactness Conditions: If* R *is a differential on a finite modular lattice* L, *then the following statements are equivalent.*

(i) R *is exact.*

(ii) *Closed Element Covering Property: If an element* y *covers an element* x *in the domain lattice* L, *then the image of* Cl(y) *is equal to or covers the image of* Cl(x) *in the closed element lattice* L/R.

(iii) *Duality:* $R^*[\tilde{y}, \tilde{x}]=1-R[x, y]$ *for all steps* [x, y].

(iv) *Existence of Dual-Closed Elements: In every interval* [x, z] *of the domain lattice on which* R *does not everywhere have the value 1, there exists a relatively dual-closed element other than the element* x.

(v) *Independence of Path: Differential sums of the form* $\sum_{i=1}^{n} R[p_{i-1}, p_i]$ *along paths* p: $p_0<p_1<\ldots<p_n$ *are dependent only upon the end points* $p_0$, $p_n$.

The closed element covering property of exact differentials, together with the construction of the closure operator $Cl$ of a differential $R$, imply that, for any step $[x, y]$ of the domain lattice $L$, $Cl(y)$ covers $Cl(x)$ in $L/R$ if and only if $R[x, y]=1$, while $Cl(y)=Cl(x)$ if and only if $R[x, y]=0$. Conversely, if $p$: $x=p_0<p_1<\ldots<p_n=z$ is any path from $x$ to $z$ in the domain lattice $L$ of an exact differential $R$, and if $Cl(z)$ covers $Cl(x)$ in the closed element lattice $L/R$, then $R[p_0, p_1]=1$ because $p_0=x$ is a closed element, and $R[p_{i-1}, p_i]=0$ for $i=2, \ldots, n$, because $Cl(p_1)$ and $Cl(p_n)$ both cover $Cl(x)$ in $L/R$, and must be equal.

Exactness of differentials is related to exactness of differential forms, in the theory of functions of several real variables. Given a differential $R$ on the Boolean algebra of a set $\{e_1, \ldots, e_n\}$, and given a subset $x$ of this set, let

$$M_i(x)=R[x, x\vee e_i]$$

be the coefficient of $de_i$ in the first order differential

$$\sigma(x)=\sum_i M_i(x)de_i.$$

Partial differentiation is defined by

$$\frac{\partial M}{\partial e_i}(x)=M(x\vee e_i)-M(x)$$

for any function $M$ from subsets to numbers. Then the usual exactness condition may be expressed in

terms of the differential $R$:

$$\frac{\partial M_i}{\partial e_j} = \frac{\partial M_j}{\partial e_i}$$

if and only if

$$R[x \vee e_j,\ x \vee e_i \vee e_j] - R[x,\ x \vee e_i]$$
$$= R[x \vee e_i,\ x \vee e_j \vee e_i] - R[x,\ x \vee e_j]$$

if and only if

$$R[x,\ x \vee e_j] + R[x \vee e_j,\ x \vee e_i \vee e_j]$$
$$= R[x,\ x \vee e_i] + R[x \vee e_i,\ x \vee e_i \vee e_j],$$

a statement which excludes inexact local graphs, yet is true on local graphs which are zero, mixed, prime, or one.

## 5. Single-Element Extensions

An understanding of the structure of the closed element lattice $L/R$ is important for the extension theorem. A lattice $Q$ is *semimodular* if, for any elements $x$, $y$ in $Q$ for which $x$ covers $x \wedge y$, it follows that $x \vee y$ covers $y$. A lattice $Q$ is *geometric* if it is semimodular and complemented, i.e., if all join-irreducible elements are atoms. In a geometric lattice, every element is expressible as a join of atoms and as a meet of coatoms.

PROPOSITION: *The Closed Element Lattice: If* R *is an exact differential on a finite modular lattice* L, *the lattice* L/R *is semimodular. If the domain lattice* L *is complemented, the lattice* L/R *is geometric. (The proof may be found in "Lattice Differentials . . ."*.)

Definition: A subset $J$ of the elements of a lattice is *convex* if $x \in J$ $z \in J$, and $x < y < z$ imply $y \in J$. A *cut* of a lattice $Q$ is a convex subset of $Q$ which, if it is nonempty, contains the lattice element 1. A cut $J$ of a lattice $Q$ is *modular* if $x \in J$, $y \in J$, and $x$ covers $x \wedge y$ imply $x \wedge y \in J$.

The concept of modular cut combines the properties of a Dedekind cut with a covering condition. A modular cut is not necessarily a lattice ideal: it is closed with respect to arbitrary multiplication ($\vee$) by any lattice element, but must be closed with respect to sums ($\wedge$) of elements in the cut only when one of the elements covers the sum. The simplest modular cuts in a lattice are the empty cut, the cut containing only the element 1, and the cut containing the entire lattice. Any ideal ($x \in J$, $y \in J$, $z \in Q$ imply $x \vee z \in J$ and $x \wedge y \in J$) is also a modular cut.

THEOREM: *Extensions Produce Modular Cuts: Let* R *be the differential of a matroid on the Boolean algebra* L *of all subsets of an* n*-element set* X: *let* e *be any element of the set* X, *and let* $R_0$ *be the restriction of the matroid* R *to the Boolean algebra* $L_0$ *of all subsets of the set* $X_0 = X - \{e\}$. *Let* J *be the set of all closed elements* x *of the matroid* $R_0$ *such that* $R[x, x \vee e] = 0$. *Then* J *is a modular cut of the lattice* $L_0/R_0$ *of closed elements of the restricted matroid.*

PROOF: If the set $J$ is nonempty, and contains some element $x$ in the closed element lattice $L_0/R_0$, we have the ordering on steps $[x, x \vee e] < [1_0, 1]$, so the unit element $1_0$ of the lattice $L_0/R_0$ is also in the set $J$. $J$ is convex, and therefore a cut, by the same argument: $x \in J$, $x \leq y$ imply $[x, x \vee e] < [y, y \vee e]$, so $R[x, x \vee e] = 0$ implies $R[y, y \vee e] = 0$ and $y \in J$.

It remains to show that the cut $J$ of the lattice $L_0/R_0$ is modular. Assume closed subsets $x$ and $y$ in $L_0$ have an intersection $z$. Assume that, as elements of the closed element lattice $L_0/R_0$, $x$ and $y$ are in the cut $J$, and $x$ covers $z$, yet $z$ is not in the cut $J$. Choose a path $p$: $z = p_0 < p_1 < \ldots < p_n = x$ from $z$ to $x$ in the Boolean algebra $L$. $R[p_{i-1}, p_i] = 1$ only for $i = 1$, because the image of $x$ covers the image of $z$ in the closed element lattice $L_0/R_0$. Since $x$ is in the cut $J$, $R[x, x \vee e] = 0$, and the sum of the values of the differential $R$ from $z$ to $x \vee e$ is 1. Lift the path $p$ to a path $p \vee e$: $z \vee e = p_0 \vee e < p_1 \vee e < \ldots < p_n \vee e = x \vee e$. Since $z$ is not in the cut $J$, $R[z, z \vee e] = 1$, and $R$ has value 0 along the entire path $p \vee e$, because the $R$-sum is independent of path.

The subset $p_1 \vee y$ covers $y$, and $R[y, p_1 \vee y] = 1$, because $y$ is closed relative to the restricted differential $R_0$. $R[y, y \vee e] = 0$, because $y$ is in the cut $J$. Thus the local graph on $y$, $p_1 \vee y$, $y \vee e$, $p_1 \vee y \vee e$ is mixed, and $R[y \vee e, p_1 \vee y \vee e] = 1$. This contradicts the projective property of the differential $R$, because $[z \vee e, p_1 \vee e] \leq [y \vee e, p_1 \vee y \vee e]$, yet $R$ has value 0 along the path $p \vee e$.

The converse of this theorem is also true, as we now prove.

THEOREM: *Modular Cuts Produce Extensions: Let* L *be the Boolean algebra of all subsets of an* n*-element set* X, *and let* e *be any element of the set* X. *Let* $R_0$ *be any matroid on the Boolean algebra* $L_0$ *of all subsets of the set* $X_0 = X - \{e\}$, *and let* J *be any modular cut of the closed element lattice* $L_0/R_0$. *There is a unique matroid* R *on the Boolean algebra* L, *extending* $R_0$ *and with the property, for all closed elements* $x \in L_0$, $R[x, x \vee e] = 0$ *if any only if* $x \in J$.

PROOF: Let $r_0$ be the Whitney rank function of the matroid $R_0$. $r_0(x) = \sum_i R_0[p_{i-1}, p_i]$ for any path $p$ from 0 to $x$ in the Boolean algebra $L_0$. Define a function $r$ on subsets in the Boolean algebra $L$ by: for every subset $x$ in $L_0$,

(i) $r(x) = r_0(x)$

(ii) $r(x \vee e) = r_0(x) + 1$ if $Cl(x) \notin J$
$= r_0(x)$ if $Cl(x) \in J$.

We prove that the function $r$ is a unit increase function. Steps in the Boolean algebra $L$ are of three types: $[x, y]$, $[x, x \vee e]$, and $[x \vee e, y \vee e]$, where $[x, y]$ is a step in the Boolean algebra $L_0$. On steps of the first type, $r$ agrees with $r_0$, a unit increase function. On steps of the second type, we have defined $r(x \vee e)$ to be

either equal to or one greater than $r(x)$. On steps of the third type, $r(y \vee e) - r(x \vee e) = r_0(y) - r_0(x)$ unless $Cl(x) \notin J$ and $Cl(y) \in J$. But if $Cl(x) \notin J$ and $Cl(y) \in J$, we know $Cl(x) < Cl(y)$, so $R_0[x, y] = 1$, and $r(x \vee e) = 1 + r_0(x) = r_0(y) = r(y \vee e)$. Thus $r$ is a unit increase function.

If $R$ is a matroid on the Boolean algebra $L$, and if $x$ is any subset of $X_0$, then $R[x, x \vee e] = R[Cl(x), Cl(x) \vee e]$, because $R$ is subadditive on steps, and has value 0 on all steps of the interval $[x, Cl(x)]$. Thus, if $R$ is a matroid on the Boolean algebra $L$, extending the matroid $R_0$ on the Boolean algebra $L_0$, and having the property, for all closed elements $x$ in $L_0$, that $R[x, x \vee e] = 0$ if and only if the image of $x$ in $L_0/R_0$ is in the modular cut $J$, then the Whitney rank function of the matroid $R$ must be the unit increase function $r$ defined above. Thus there exists at most one matroid $R$ with the required properties. If the unit increase function $r$ can be shown to be a Whitney rank function, the existence of such a matroid $R$ extending $R_0$ is established.

Define the function $R$ on all steps of the Boolean algebra $L$ to be the difference of the unit increase function $r$. We must show that no local graph of $R$ on $L$ is inverted. The local graphs in $L$ are of three types:

(i) Local graphs on subsets $x \wedge y$, $x$, $y$, $x \vee y$, where neither $x$ nor $y$ contain the element $e$, but both cover $x \wedge y$.

(ii) Local graphs on subsets $x$, $y$, $x \vee e$, $y \vee e$, where $y$ covers $x$ and does not contain $e$.

(iii) Local graphs on subsets $z$, $x \vee z$, $y \vee z$, $x \vee y \vee z$, where $z = (x \wedge y) \vee e$ and the local graph on $x \wedge y$, $x$, $y$, $x \vee y$ is of type (i). No local graph of type (i) is inverted because $R_0$ is exact. No local graph of type (ii) is inverted, because $R[x, x \vee e] = 0$ implies $Cl(x) \in J$, and $x < y$ implies $Cl(y) \in J$ and $R[y, y \vee e] = 0$.

If a local graph of type (iii) is inverted, the values of the difference $R$ are uniquely determined on the interval $[x \wedge y, x \vee y \vee z]$ of length 3. The local graph on $x \wedge y$, $x$, $y$, $x \vee y$ is one, by the projective property of $R$ across two local graphs of types (ii) and one of type (i), which are known to be either zero, mixed, prime, or one. The local graphs of type (ii) on $x \wedge y$, $y$, $z$, $y \vee z$ and on $x \wedge y$, $x$, $z$, $x \vee z$ are then prime, and that on $x$, $x \vee y$, $x \vee z$, $x \vee y \vee z$ is mixed. The resulting configuration is drawn in figure 3. This configuration is excluded by the modularity of the cut $J$, because the closures of $x$ and of $y$ cover the closure of $x \wedge y$ and are covered by the closure of $x \vee y$ in the closed element lattice $L_0/R_0$, while $R[x, x \vee z] = R[y, y \vee z] = 0$ implies $Cl(x)$ and $Cl(y)$ are in the modular cut $J$, and $R[x \wedge y, z] = 1$ implies $Cl(x \wedge y)$ is not in $J$.

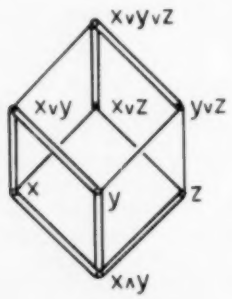



FIGURE 3

## 6. Linear Subclasses

A preponderance of matroid theory is phrased in terms of circuits or bonds. Systematically, circuits and bonds are classes of distinguished subsets of a set, satisfying two axioms. In terms of closures, circuits are minimal nonempty dual-closed subsets. The circuits of the dual matroid (the bonds of the matroid) appear naturally as the set-complements of those closed subsets which are coatoms of the closed-set lattice.

Single-element extensions of matroids are readily characterized in terms of coatoms of the closed-set lattice. We show in this section how modular cuts are generated by linear subclasses [5] of the set of coatomic closed subsets.

The passage from modular cuts to linear subclasses may be accomplished within the closed-set lattice. No reference is made to the manner in which this lattice is embedded in the domain of the matroid. Closed-set lattices of matroids are geometric: the structure differential of a finite geometric lattice is a matroid. The following theorems may thus be regarded as theorems about finite geometric lattices, and will be so written.

A *linear subclass* of the set $C$ of coatoms of a geometric lattice $Q$ is any subset $\mathcal{L}$ of the set $C$ which, if it contains two coatoms which cover their infimum, contains all other coatoms covering that same infimum.

The *convex closure* of a set $\mathcal{L}$ of coatoms of a geometric lattice $Q$ is the set of all lattice elements $x$ such that they and all lattice elements $y > x$ are expressible as an infimum of coatoms in the set $\mathcal{L}$.

THEOREM: *Generation of Modular Cuts: The convex closure of any subset $\mathcal{L}$ of the set of coatomic elements of a finite geometric lattice* Q *is a modular cut of the lattice* Q *if and only if the set $\mathcal{L}$ is a linear subclass. Every modular cut is the convex closure of the set of coatomic elements it contains.*

PROOF: Let $J$ be a nonempty modular cut of a finite geometric lattice $Q$. Every element of $J$ is expressible as an infimum of coatomic elements of $Q$, because $Q$ is geometric. All coatomic elements in such an expression are in the modular cut $J$, because $J$ is convex and contains the element 1. Thus $J$ is contained in the convex closure of the set of coatomic elements in $J$. If an element $x$ of the lattice $Q$ is in the convex closure of the set of coatomic elements in $J$, choose a path $p$: $x = p_0 < p_1 < \ldots < p_n = 1$ from $x$ to 1. Since $p_{n-1}$ is expressible as an infimum of coatomic elements in $J$, yet is a coatomic element, $p_{n-1}$ is an element of $J$. Let $j$ be the largest index such that $p_j \notin J$, then find a coatomic element $z$ in $J$ such that $p_{j+1} \wedge z = p_j$. Since

[5] W. T. Tutte, Trans. A.M.S., 1958.

the cut $J$ is modular, $p_j$ is also in $J$. Thus the element $x$ is in the modular cut $J$, and every modular cut is the convex closure of the set of coatomic elements it contains.

Assume again that $J$ is a modular cut of a finite geometric lattice $Q$, and let $L$ be the set of coatomic elements in $J$. If $x$, $y$, and $z$ are coatomic elements of $Q$ such that $x$ and $y$ cover $x \wedge y$ and such that $x$ and $y$ are in the set $L$, then $x \wedge y$ is in the modular cut $J$. Since $J$ is convex, $z$ is also in $J$. Being coatomic, $z$ is in $\mathscr{L}$, and $\mathscr{L}$ is a linear subclass.

It remains to prove that the convex closure of any linear subclass is a modular cut. One step in this proof is of independent interest, and is set apart as a lemma.

LEMMA: *Covering Properties of Linear Subclasses: If $\mathscr{L}$ is a linear subclass of the coatomic elements of a finite geometric lattice* Q, *and if* $C(\mathscr{L})$ *is the convex closure of $\mathscr{L}$, then every lattice element* $w \notin C(\mathscr{L})$ *is covered by at most one element of* $C(\mathscr{L})$.

PROOF: Let $w$ be a lattice element not in $C(\mathscr{L})$, yet covered by two distinct elements $u_1$ and $u_2$ of $C(\mathscr{L})$. Assume $w$ is maximal with respect to these properties. Since every element $x$ above $w$ is an infimum of coatomic elements above $w$, yet some element above $w$ is not an infimum of coatomic elements in the linear subclass $\mathscr{L}$, we may choose a coatomic element $z$, $w \leq z$, $z \notin L$. The rank $r(w)$ of the element $w$ in the lattice $Q$ is equal to neither $r(1)$ nor $r(1)-1$, because $w$ is covered by two distinct lattice elements. If $r(w)=r(1)-2$, $w$ is covered by two coatoms in the linear subclass $\mathscr{L}$, and thus must be in $C(\mathscr{L})$. Thus $r(w) \leq r(1)-3$.

Choose a path $p$: $w=p_0 < p_1 < \ldots < p_k=z$ from $w$ to $z$, and let $q$ be any complement ($p_1 \vee q=p_2$ and $p_1 \wedge q=w$) of $p_1$ in the lattice interval $[w, p_2]$ of height 2. Note that no elements $p_i$ of the path $p$ are in $C(\mathscr{L})$, nor is $q$ in $C(\mathscr{L})$. We shall prove that either $p_1$ or $q$ is also covered by distinct elements of the convex closure $C(\mathscr{L})$, in contradiction to the maximality of $w$.

The elements $p_1 \vee u_1$, $p_1 \vee u_2$, and $u_1 \vee u_2$ are either all distinct or all equal; they are all in $C(\mathscr{L})$ because they lie above either $u_1$ or $u_2$; none of them is equal to $p_2$. If they are distinct, $p_1$ is covered by $p_1 \vee u_1$ and $p_1 \vee u_2$, both in $C(\mathscr{L})$. If they are all equal, they are not equal to $p_2$, so the elements $q \vee p_1$, $q \vee u_1$, $q \vee u_2$ are all distinct. But $q \vee u_1$ and $q \vee u_2$ are in $C(\mathscr{L})$, and cover $q$.

The remainder of the proof of the theorem on generation of modular cuts now follows. Let $\mathscr{L}$ be a linear subclass of the set of coatoms of a finite geometric lattice $Q$, and let $C(\mathscr{L})$ be the convex closure of $\mathscr{L}$. The element 1 is in $C(\mathscr{L})$, because 1 is the infimum of the empty subset of $\mathscr{L}$. Convexity of $C(\mathscr{L})$ follows directly from the definition of convex closure, so $C(\mathscr{L})$ is a cut of the lattice $Q$. Assume $x$ and $y$ are in $C(\mathscr{L})$, and that $x$ covers $x \wedge y$. Choose a path $p$: $x \wedge y=p_0 < p_1 < \ldots < p_k=y$ from $x \wedge y$ to $y$. Since $x$ covers $x \wedge y$, and $x \wedge y=x \wedge p_i$ ($i=0, 1, \ldots, k$), $x \vee p_i$ covers $p_i (i=0, 1, \ldots, k)$. The elements $x \vee p_i$ are in $C(\mathscr{L})$, because $x \leq x \vee p_i$. Let $j$ be the least index such that $p_j \in C(\mathscr{L})$. If $j \neq 0$, $p_{j-1}$ is covered by both $x \vee p_{j-1}$ and $p_j$, both of which are in $C(\mathscr{L})$. Applying the lemma on covering properties of linear subclasses, we see $p_0=x \wedge y$ is in $C(\mathscr{L})$, verifying the modular property of the cut $C(\mathscr{L})$.

The characterization of modular cuts as convex closures of linear subclasses provides us with a description of single element extensions of matroids. If $R$ is a matroid defined on subsets of a set $X=X_0 \cup \{e\}$, extending a matroid $R_0$ defined on subsets of the set $X_0$, then the element $e$ is dependent upon a subset $x$, i.e., $R[x, x \vee e]=0$, if and only if $Cl(x)$ is in the convex closure of some linear subclass of the coatoms of $L_0/R_0$.

Let us now characterize the coatomic closed subsets of the extended matroid. We aim for a simple construction of the coatoms of an extended matroid, applicable to matrix-theoretic investigations.

THEOREM: *Closed Subsets of Single-Element Extensions: Let* R *be a matroid on the Boolean algebra* L *of all subsets of a finite set* $X=X_0 \cup \{e\}$, *which extends a matroid* $R_0$ *on the Boolean algebra* $L_0$ *of all subsets of the set* $X_0$, *and which is produced by the modular cut* J *of the lattice* $L_0/R_0$. *Then the closed subsets of the matroid* R *are*

(i) *Closed subsets relative to* $R_0$ *which are not in the modular cut* J,

(ii) *Subsets of the form* $x \vee e$ *where* x *is a closed subset relative to* $R_0$ *in the modular cut* J,

(iii) *Subsets of the form* $x \vee e$ *where* x *is a closed subset relative to* $R_0$ *not in the modular cut* J, *and covered in the lattice* $L_0/R_0$ *only by elements not in the modular cut* J.

PROOF: Subsets of $X$ not containing the element $e$ are closed relative to $R$ only if they are closed relative to $R_0$, and are not in the modular cut $J$. Subsets containing the element $e$ and closed relative to $R$ must, on deletion of $e$, become closed elements relative to $R_0$, because $R$ is projective on steps. To ascertain which subsets of the form $x \vee e$, for subsets $x$ closed relative to $R_0$, are closed relative to $R$, we observe that the $R$ value 1 on a step $[x, y]$ for $y \in L_0$ decreases to 0 on the step $[x \vee e, y \vee e]$ if and only if the local graph on $x$, $y$, $x \vee e$, $y \vee e$ is prime. But this is true if and only if $Cl(y)$, an element in the modular cut $J$, covers (in $L_0/R_0$) $Cl(x)$, an element not in $J$. The subset $x \vee e$ is closed relative to $R$ if and only if this occurs for no subset $y$ covering $x$.

The coatomic closed subsets of the extended matroid are easily picked out as those closed subsets $x$ such that paths $p$ from $x$ to 1 have $R$ value 1 only on the first step. They are those coatomic closed subsets of $X_0$ which are not in the linear subclass, together with subsets of the form $x \vee e$, $x$ being either a coatomic closed subset in the linear subclass, or a subset of rank $r(1)-2$ and covered in $L_0/R_0$ only by coatomic closed subsets not in the linear subclass.

Taking set-complements of coatomic closed subsets, we obtain a statement of the single-element extension theorem in terms of the bonds or circuits of a matroid. A linear subclass of a set of bonds is a set of bonds which contains, along with any pair $x$, $y$ of bonds whose union $x \cup y$ has rank 2 in the lattice

of unions of bonds, any other bonds contained in the union $x \cup y$. The word "circuit" may be substituted everywhere for "bond" in the previous sentence.

THEOREM: *Bonds of Single-Element Extensions: Given a matroid with differential* $R_0$ *on the Boolean algebra of subsets of a finite set* $X_0$, *and given a linear subclass* $\mathcal{L}$ *of the set* $B_0$ *of bonds of* $R_0$, *let* X *be the set* $X_0 \cup \{e\}$. *Then the following are the bonds of the matroid* R *produced by the linear subclass* $\mathcal{L}$:

(i) *Bonds of* $R_0$ *in the linear subclass* $\mathcal{L}$;

(ii) *Sets of the form* $x \cup \{e\}$, *where* x *is a bond of* $R_0$ *outside the linear subclass* $\mathcal{L}$;

(iii) *Unions of bonds of* $R_0$, *of rank 2 in the lattice of unions of bonds, and covering no bond in the linear subclass* $\mathcal{L}$.

As a simple example, consider the bond matroid of a triangle with sides a, b, c, and an isthmus d attached at one of the vertices. The bonds are ab, ac, bc, and d. The single element extension corresponding to the linear subclass {ab} has bonds (i) ab, (ii) ace, bce, de, (iii) acd, bcd. These are the bonds of the square with edges a, b, e, d and a diagonal c making a triangle with a and b. The original graph is obtained by deleting the edge e.

In the above theorem and example, a matroid $R_0$ is extended to a matroid $R$ in such a way that $R_0$ is the restriction of $R$ to the Boolean algebra interval $[0, 1-e]$. This restriction is Tutte's[6] $R_0 = R \cdot X_0$, and is accomplished by "elimination" of the element $e$. Applying the above theorem to the dual matroid $R^*$ and interpreting the results in terms of the matroid $R$, we obtain a characterization of matroids $R$ such that $R_0$ is the restriction of $R$ to the Boolean algebra interval $[e, 1]$. This restriction is Tutte's $R_0 = R \times X_0$, and is accomplished by "contraction" of the element $e$. Such extensions are produced by linear subclasses of the set $C_0$ of circuits of the matroid $R_0$; the circuits of the extended matroid are described in the above theorem, if "circuit" is substituted for "bond" throughout.

A class of unsolved problems suggested by Rota and Tutte is the characterization of those linear subclasses or modular cuts on matroids with property $P$, for which the associated single-element extension also has property $P$. This property $P$ may be binary, regular, even, graphic, or any property of the lattice of closed subsets. Characterizations of linear subclasses or modular cuts which preserve these properties may simplify inductive proofs of otherwise difficult theorems.

## 7. Geometric Lattices

The analysis of closed subsets of single-element extensions completed in the previous section, provides a method for the construction of all finite geometric lattices. Since every finite geometric lattice is the closed element of its structure differential, and since all matroids may be constructed inductively by modular cuts, all finite geometric lattices are constructed by a sequence of modular cuts beginning with the simplest geometric lattice: that containing only the elements 0 and 1.

Referring to the theorem on closed subsets of single-element extensions, we see that the subsets of the first two types (those closed subsets $x \notin J$ and those closed subsets $x \vee e$ with $x \in J$) form a subsystem isomorphic to the closed-set lattice of the matroid $R_0$. Subsets of the third type must be added to the closed-set lattice, and the appropriate covering lines must be added to the lattice diagram, as follows.

Given a finite geometric lattice $Q$ and any modular cut $J$ of $Q$, let *Ext* $J$ be composed of those elements not in $J$, and covered by no elements of $J$. A new lattice $Q'$ may be defined, extending $Q$, by adding to the lattice $Q$ an element $x'$ corresponding to each element $x$ in *Ext* $J$, and by extending the covering relation "$y$ covers $x$" to include elements $x$ and $y$ of $Q'$ such that

(i) $x, y$ are in $Q$, and $y$ covers $x$ in $Q$

(ii) $y = x'$

(iii) $x = z'$, $y \in J$, and $[z, y]$ is an interval of length 2 in the lattice $Q$.

This inductive procedure is carried out for four steps in figure 4.

Although any modular cut yields an extension of the structure differential of a geometric lattice $Q$, only those modular cuts $J$ for which *Ext* $J$ is non-empty change the lattice $Q$. The set *Ext* $J$ is non-empty, and thus the extension is nontrivial on $Q$, if and only if no atom of the lattice $Q$ is in the modular cut $J$.

## 8. Numerical Classification of Matroids

We have seen the application of the single-element extension theorem to the construction of finite geometric lattices. Let us now turn our attention to a numerical classification of matroids. We prove the existence, on any $n$-element set, of at least one matroid in each of $2^n$ classes, and thereby obtain a lower bound on the number of different matroids on a given set.

The sequence of values of the differential of a matroid on the steps of a path from 0 to 1 in the domain lattice may be thought of as forming a word in a language employing two letters, the letter 0 coming before the letter 1 in the alphabet. Of the words thus associated with a given matroid the word coming first in alphabetical order is an isomorphy invariant of the matroid, and serves as an index for a classification system.

We use the term *least path* to indicate a path, from 0 to 1 in the domain lattice of a matroid, if the word formed of the differential values along this path is alphabetically the first among all such words associated with the same matroid. The word associated with a least path of a matroid we shall call the *first word* of the matroid.

Given any $n$-letter word formed of the letters 0 and 1, we shall prove the existence of a matroid having that word as first word. This establishes the exist-

[6] W. T. Tutte, Trans. A.M.S., 1958.

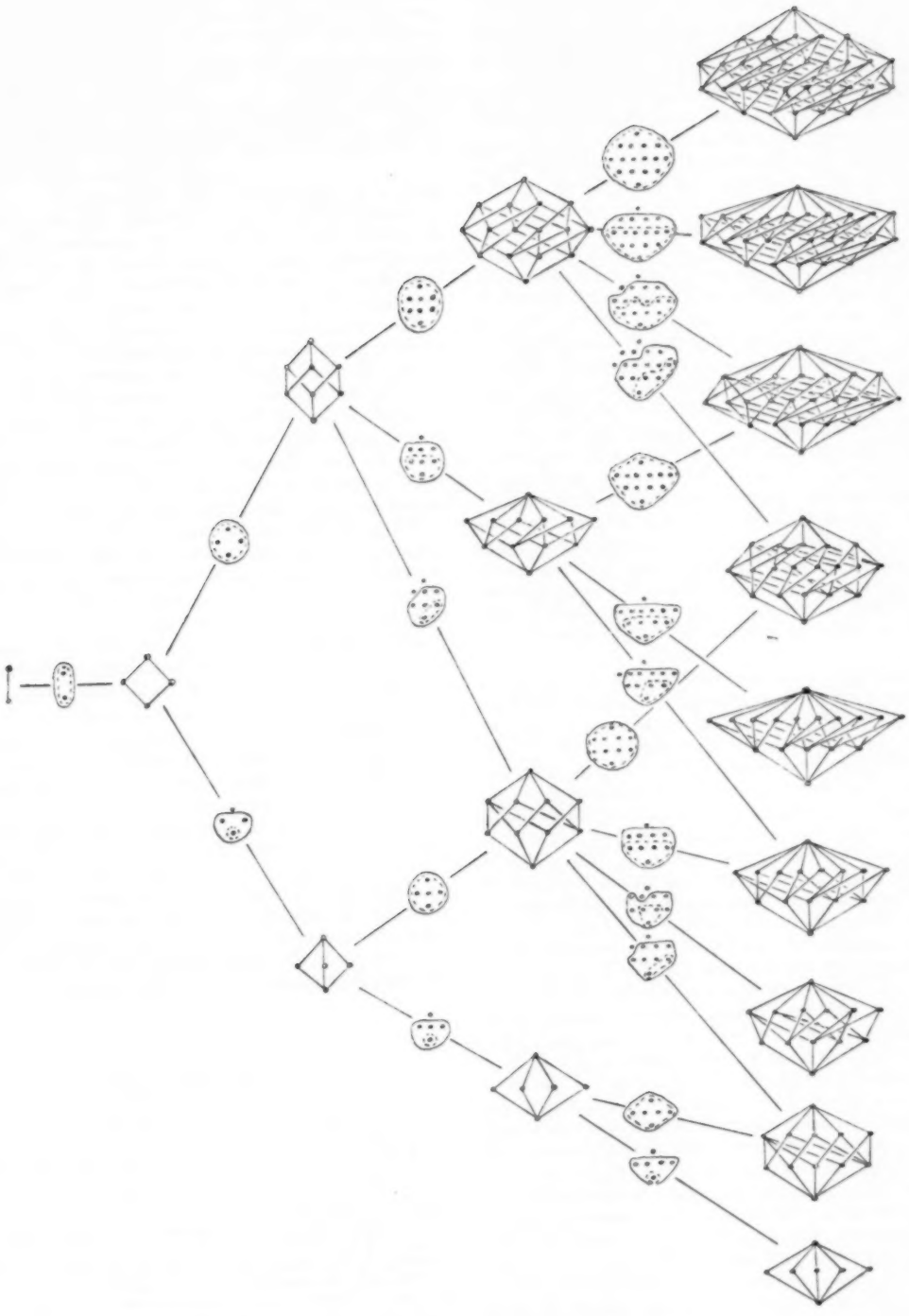


FIGURE 4. *Geometric lattices with* n *atoms,* n=*1, 2, 3, 4, 5.*

Modular cuts generating each extension are indicated on the connecting lines. The set $'f$ is marked with a solid line; the set Ext $f$ with a dotted line.



757-615 O-65-5

ence of at least $2^n$ isomorphically inequivalent matroids defined on the Boolean algebra of subsets of an $n$-element set. The existence proof proceeds by induction, and utilizes our knowledge of single-element extensions.

THEOREM: *Existence of Matroids With A Given First Word: Let* W *be any* n*-letter word, i.e., any sequence of length* n *consisting of zeros and ones. There exists a matroid* R *on a Boolean algebra* L *of all subsets of an* n*-element set, such that the word* W *is the first word of the matroid* R *on* L.

PROOF: The theorem is obvious for $n=1$. Assume we are given a word $W=w_1, \ldots, w_n$ of length $n$, and that for any word of length $n-1$ there exists a matroid defined on a Boolean algebra of all subsets of an $(n-1)$-element set, such that the word of length $n-1$ is the first word. If $w_n=1$, find a matroid $R_0$ on the interval $[0, 1-e]$, for which $w_1, \ldots, w_{n-1}$ is first word. Choose the empty modular cut, and let $R$ be the matroid produced by that cut. Then $R[x, x\vee e]=1$ for all subsets $x$ in the lattice interval $[0, 1-e]$, and $R[x\vee e, y\vee e]=R[x, y]$ for all pairs $x$, $y$ of elements in the lattice interval $[0, 1-e]$. That $W$ is the first word for $R$ follows from the fact that the word for a path employing a step greater than $[0, e]$ in any position but the last may be obtained from the corresponding word for the projection of this path into the sublattice $[0, 1-e]$ by insertion of the letter 1 at some point, moving all later letters back one space. This results in a word at least as late as $W$ in the alphabetical ordering.

If the final letter $w_n$ in the word $W$ is 0, the proof is a bit more intricate. Form a matroid $R_0$ on the Boolean algebra interval $[0, 1-e]$, having $w_1, \ldots, w_{n-1}$ as first word. On the lattice of closed subsets $[0, 1-e]/R_0$, let a modular cut $J$ contain only the closed subset $1-e$ of $R_0$. Construct an extended matroid $R$ on the Boolean algebra $L$ of all subsets of the $n$-element set in accordance with the extension theorem. The coatomic closed subsets of the resulting differential $R$ are the coatomic closed subsets of $R_0$ on the interval $[0, 1-e]$, together with those subsets of the form $x\vee e$, where $x$ is a closed element of $L_0/R_0$ of rank $r(1)-2$, because no coatoms of $L_0/R_0$ are in the modular cut $J$.

If a subset $x$ in the interval $[0, 1-e]$ is not connected to $1-e$ by a path along which $R_0=0$, it is contained in some coatomic closed subset of $R_0$ not in the modular cut $J$, so $R[x, x\vee e]=1$. In any path from 0 to $x\vee e$ for any such set $x$, the projection of this path into the lattice interval $[0, 1-e]$, then via the step $[x, x\vee e]$, is a lesser path. Let $p$ be a least path for the extended matroid $R$, and one which passes through a minimum number of subsets containing the element $e$. By the argument just given, if the subset $p_k$ is the first subset containing the element $e$ occurring in the path $p$, the subset $p_{k-1}$ either must be a coatomic closed subset relative to the matroid $R_0$, or must be connected to $1-e$ by a path along which $R_0=0$. If the subset $p_{k-1}$ is connected to $1-e$ by a path along which $R_0=0$, all steps above $p_{k-1}$ in the path $p$ have $R$ value zero, so the path following $p$ to $p_{k-1}$, then via $1-e$ to 1 gives rise to the same word, and involves the element $e$ in fewer subsets. On the other hand, if $p_{k-1}$ is a coatomic closed subset of $R_0$, the step $[p_{k-1}, p_k]$ is the final step of $p$ for which $R$ has value one. Any path along $p$ to $p_{k-1}$, then via $1-e$ to 1 gives rise to the same word, and involves the element $e$ in fewer subsets. Thus the path $p$ passes through $1-e$, and is a least path. The restriction of $p$ to the interval $[0, 1-e]$ is a least path for $R_0$, so must give rise to the word $w_1, \ldots, w_{n-1}$. Since $w_n$ is 0, the word for the least path $p$ of the matroid $R$ is the word $W$.

COROLLARY: *A Lower Bound for the Number of Matroids: There are at least* $2^n$ *nonisomorphic matroids on a Boolean algebra of all subsets of an* n*-element set.*

PROOF: There are $2^n$ different words of length $n$, and each is the first word for some matroid.

The construction of a matroid with a given first word may be carried out methodically. On a diagram of the Boolean algebra of all subsets of an $n$-element set, choose any path to be the least path, and indicate on it the letters of the given first word. Then mark all steps according to the projective and subadditive properties of differentials, the requirement that every local graph be zero, mixed, prime or one, and the requirement that the given path be least.

When all the implications of the first word are exhausted, a matroid may not be fully determined. A matroid can then be defined in more than one way with the given first word: the number of such ways we shall term the multiplicity of the word.

DEFINITION: The *multiplicity* $\theta(W)$ of a word $W$ of length $n$ is the number of isomorphically inequivalent matroids with first word $W$, definable on the Boolean algebra of all subsets of an $n$-element set.

We have proven that all words have multiplicity $\theta$ at least equal to one. All words of one, two, or three letters have multiplicity equal to one. A single four letter word, 1010, has multiplicity two.

It is interesting to speculate about what additional isomorphy invariants may serve to classify matroids. The set of matroids with a given first word may be arranged in a partially ordered system, along paths of which the *complexity* [7] (the number of bases in the matroid) increases. This investigation, which is involved with the dichromate (see footnote 7) of a matroid, is the subject of a future paper.

## 9. Bibliography


Birkhoff, G. (1935), Abstract linear dependence and lattices, Amer. J. **57**, 800–804.

Crapo, H. H., Lattice differentials, and the theory of combinatorial independence. (to appear).

Dilworth, R. P. (1941), The arithmetical theory of Birkhoff lattices, Duke J. **8**, 286–299.

——— (1944), Dependence relations in a semi-modular lattice, Duke J. **11**, 575–587.

Edmonds, J., et al. (Sept. 1964), Seminar: The Mathematical Theory of Matroids, National Bureau of Standards.

MacLane, S. (1938), A lattice formulation for transcendence degrees and $p$-bases, Duke J. **4**, 455–468.


[7] W. T. Tutte, Can. J., 1954.

Ore, O. (1962), Theory of Graphs, A.M.S. Colloquium Publications **38.**
Rado, R. (1942), A Theorem on independence functions, Quart. J. of Math. (Oxford) **13,** 83–89.
——— Note on Independence Functions, Proc. Lond. Math. Soc. (3) **7,** 300–320 (1957).
Rota, G.-C. (1964), On the foundations of combinatorial theory, Zeitschr. fur Wahrscheinlichkeitstheorie **2,** 340–368.
Szasz, G., Introduction to Lattice Theory (Academic Press, New York, 1963).
Tutte, W. T. (1954), A Contribution to the theory of chromatic polynomials, Can. J. **6,** 80–91.
——— A Class of Abelian Groups, Can. J. **8,** 13–28 (1956).
——— A homotopy theorem for matroids, I-II, Trans. A.M.S. **88,** 144–174 (1958).
——— Matroids and Graphs, Trans. A.M.S. **90,** 527–552 (1959).
Whitney, H. (1935), On the abstract properties of linear dependence, Amer. J., **57,** 509–533.

(Paper 69B1–133)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Minimum Partition of a Matroid Into Independent Subsets[1]

Jack Edmonds

(December 1, 1964)

A matroid $M$ is a finite set $M$ of elements with a family of subsets, called independent, such that (1) every subset of an independent set is independent, and (2) for every subset $A$ of $M$, all maximal independent subsets of $A$ have the same cardinality, called the rank $r(A)$ of $A$. It is proved that a matroid can be partitioned into as few as $k$ sets, each independent, if and only if every subset $A$ has cardinality at most $k \cdot r(A)$.

## 1.0. Introduction

Matroids can be regarded as a certain abstraction of matrices [8].[2] They represent the properties of matrices which are invariant under elementary row operations but which *are not* invariant under elementary column operations—namely properties of dependence among the columns. For any matrix over any field, there is a matroid whose elements correspond to the columns of the matrix and whose independent sets of elements correspond to the linearly independent sets of columns. A matroid $M$ is completely determined by its elements and its independent sets of elements.

The same letter will be used to denote a matroid and its set of elements. The letter $I$ with various sub or superscripts will be used to denote an independent set.

The interest of matroids does not lie only in how they generalize some known theorems of linear algebra. There are examples, which I shall report elsewhere, of matroids which do not arise from any matrix *over any field*—so matroid theory does truly generalize an aspect of matrices. However, matroid theory is justified by new problems in matrix theory itself—in fact by problems in the special matrix theory of graphs (networks). It happens that an axiomatic matroid setting is most natural for viewing these problems and that matrix machinery is clumsy and superfluous for viewing them. The situation is somewhat similar to the superfluity of (real) matrices to the theory of linear operators, though there a quite different aspect of matrices is superfluous. When it comes to implementing either theory, matrices are often the way to do it.

Matroid theory so far has been motivated mainly by graphs, a special class of matrices. A graph $G$ may be regarded as a matrix $N(G)$ of zeroes and ones, mod 2, which has exactly two ones in each column. The columns are the edges of the graph and the rows are the nodes of the graph. An edge and a node are said to meet if there is a one located in that column and that row. Of course a graph can also be regarded visually as a geometric network. It is often helpful to visualize statements on matroids for the case of graphs, though it can be misleading. Matroids do not contain objects corresponding to nodes or rows.

Theorem 1 on "minimum partitions," the subject of this paper, was discovered in the process of unifying results described in the next paper, "On Lehman's Switching Game and a Theorem of Tutte and Nash-Williams" (denoted here as "Part II"), which is a direct sequel. Theorem 1 is shown there to be closely related to those results. Lately, I have learned that Theorem 1 for the case of graphs (see sec. 1.7) was anticipated by Nash-Williams [5].

By borrowing from work of others, I intend that this paper together with possible sequels be partly expository and technically almost self-contained.

### 1.1. The Problem

Various aspects of matroids—in particular, the first pair of axioms we cite—hold intrinsic interest which is quite separate from linear algebra.

AXIOM 1: *Every subset of an independent set of elements is independent.*

Any finite collection of elements and family of so-called independent sets of these elements which satisfies axiom 1 we shall call an independence system. This also happens to be the definition of an abstract simplicial complex, though the topology of complexes will not concern us.

It is easy to describe implicitly large independence systems which are apparently very unwieldy to analyze. First example: given a graph $G$, define an *independent* set of nodes in $G$ to be such that no edge of $G$ meets two nodes of the set. Second example: define an independent set of edges in $G$ to be such that

[1] Sponsored by the Army Research Office (Durham). Presented at the Seminar on Matroids, National Bureau of Standards, Aug. 31-Sept. 11, 1964. I am much indebted to Alfred Lehman for encouraging my interest in the subject.

[2] Figures in brackets indicate the references at the end of this paper.

no node meets two edges of the set. Third example: define an independent set of edges in $G$ to be such that the edges of the set, as column vectors of $N(G)$, are linearly independent. The third example is the prototype of the systems we shall study here.

A *minimum coloring* of the nodes of a graph $G$ is a partition of the nodes into as few sets (colors) as possible so that each set is independent. A good characterization of the minimum colorings of the nodes in a graph is unknown (unless the graph is bipartite, i.e., the nodes can be colored with two colors). To find one would undoubtedly settle the "four color" conjecture.

A problem closely related to minimum coloring is the "packing problem." That is to find a good characterization (and an algorithm) for maximum cardinality independent sets. More generally the "weighted packing problem" is, where each element of the system carries a real numerical weight, to characterize the independent sets whose weight-sums are maximum. The packing problem for the systems of the first example is also very much unsolved (unless the graph is bipartite).

The minimum coloring problem for the systems of the second example is unsolved (unless the graph is bipartite). Its solution would also undoubtedly settle the four-color conjecture. However the packing problem, and more generally the weighted packing problem, is solved for the second example by the extensive theory of "matchings in graphs."

For the third example the packing problem is in a sense trivial. It is well known that the system of linearly independent sets of edges in a graph, and more generally the system of linearly independent sets of columns in a matrix, satisfies the following:

AXIOM 2: *For any subset $A$ of the elements, all maximal independent sets contained in $A$ contain the same number of elements.*

A *matroid* is a (finite) system of elements and sets of elements which satisfies axioms 1 and 2.

For any independence system, any *subsystem* consisting of a subset $A$ of the elements and all of the independent sets contained in $A$ is an independence system. Thus, a matroid is an independence system where the packing problem is postulated to be trivial for the system and all of its subsystems. For me, having spent much labor on packing problems, it is pleasant to study such systems. Matroids have a surprising richness of structure, as even the special case of graphic matroids shows.

Clearly, a subsystem of a matroid $M$ is a matroid. We call it a *submatroid* and we use the same symbol to denote it and its set of elements. The *rank*, $r(A)$, of a set $A$ of elements in $M$ or the *rank*, $r(A)$, of the submatroid $A$ of $M$ is the number of elements in each maximal independent set contained in $A$, i.e., the number of elements in a base of $A$.

The main result of this paper is a solution of the minimum coloring problem for the independent sets of a matroid. Another paper will treat the weighted packing problem for matroids.

## 1.2. Ground Rules

One is tempted to surmise that a minimum coloring can be effected for a system by some simple process like extracting a maximal independent set to take on the first color, then extracting a maximal independent set of what is left to take on the second color, and so on till all elements are colored. This is usually far from being successful even for matroids, though it is precisely matroids for which a similar sort of monotonic procedure always yields a maximum cardinality independent set and, as we shall see, in another paper, also always yields a maximum weight-sum independent set when the elements carry arbitrary real weights.

Consider the class of matroids implicit in the class $M_F$ of all matrices over fields of integers modulo primes. (For large enough prime, this class includes the matroid of any matrix over the rational field.) We seek a good algorithm for partitioning the columns (elements of the matroid) of any one of the matrices (matroids) into as few sets as possible so that each set is independent. Of course, by carrying out the monotonic coloring procedure described above in all possible ways for a given matrix, one can be assured of encountering such a partition for the matrix, but this would entail a horrendous amount of work. We seek an algorithm for which the work involved increases only algebraically with the size of the matrix to which it is applied, where we regard the size of a matrix as increasing only linearly with the number of columns, the number of rows, and the characteristic of the field. As in most combinatorial problems, finding a finite algorithm is trivial but finding an algorithm which meets this condition for practical feasibility is not trivial.

We seek a good characterization of the minimum number of independent sets into which the columns of a matrix of $M_F$ can be partitioned. As the criterion of "good" for the characterization we apply the "principle of the absolute supervisor." The good characterization will describe certain information about the matrix which the supervisor can require his assistant to search out along with a minimum partition and which the supervisor can then use with ease to verify with mathematical certainty that the partition is indeed minimum. Having a good characterization does not mean necessarily that there is a good algorithm. The assistant might have to kill himself with work to find the information and the partition.

Theorem 1 on partitioning matroids provides the good characterization in the case of matrices of $M_F$. The proof of the theorem yields a good algorithm in the case of matrices of $M_F$. (We will not elaborate on how.) The theorem and the proof apply as well to all matroids via the matroid axioms. However, the "goodness" for matrices depends on being able to carry out constructively with ease those matrix operations which correspond to the existential assertions of the theory. A fundamental problem of matroid theory is to find a good representation for general matroids—good perhaps relative to the rank and the number of elements in the matroids. There is a very

elegant lattice representation (geometric lattices, [1, 2]), but it is not something you would want to record except for the very simplest matroids.

### 1.3. The Theorem

The cardinality of a set $A$ is denoted by $|A|$. The rank of a set $A$ is denoted by $r(A)$.

THEOREM 1: *The elements of a matroid* M *can be partitioned into as few as* k *sets, each of which is independent, if and only if there is no subset* A *of elements of* M *for which*

$$|A| > k \cdot r(A).$$

The theorem makes sense for any independence system $M$ if we define the rank $r(A)$ of any subset $A$ to be the maximum cardinality of an independent set in $A$. In fact, the "only if" part of the theorem is true for any independence system $M$. Let $I_i (i = 1, \ldots, k)$ be $k$ independent sets in $M$ for which

$$\bigcup_{i=1}^{k} I_i = M.$$

For any subset $A$ of $M$, $|I_i \cap A| \leq r(A)$ and

$$|A| \leq \sum_{i=1}^{k} |I_i \cap A| \leq k \cdot r(A).$$

Thus the "only if" part is proved.

In general for the coloring problem in nonmatroidal systems, the other half of the theorem is not true. However, the Konig theorem on matchings in bipartite graphs can be regarded as a valid instance of theorem 1 for certain nonmatroidal systems. A bipartite graph is a graph whose nodes can be partitioned into two sets each independent (by coincidence, an instance of the coloring problem in our first example). The Konig theorem says that for a bipartite graph $G$ the minimum number of nodes which meet all the edges equals the maximum number of edges such that no node meets more than one of them. (This theorem solves the packing problem for a special case of our second example of independence system.)

Fourth example: For a graph $G$, let the elements of the system $M$ be the edges of $G$. For each node of $G$, let the set of edges which meet the node be an independent set in $M$. Let the subsets of these sets be the rest of the independent sets in $M$. The Konig relation for a graph $G$ implies theorem 1 for system $M$.

Theorem 1 for the system $M$ arising from $G$ does not imply the Konig theorem for $G$. For independence systems in general the relation represented by theorem 1 is weaker than the relation represented by the Konig theorem—the latter being that the minimum number of independent sets which together contain all the elements equals the maximum number of elements in a set of rank one. It's nice to have the weaker relation of theorem 1 because it might apply to other systems where the well known Konig relation does not.

### 1.4. Terminology

There are various families, (1) through (6), of subsets of the elements in a matroid $M$ which are used in describing the structure of $M$.

(1) The family of independent sets of $M$.

(2) The family of minimal dependent sets of elements in $M$ (where dependent means not independent). These are called the *circuits* in $M$. The letter $C$ with various sub or superscripts will be used to denote a circuit.

(3) The family of *spans* or *closed sets* in $M$. A *span* $S$ in $M$ is a set of elements such that no circuit of $M$ contains exactly one element not in $S$. That is, $|S \cap C| \neq 1$ for every circuit $C$ in $M$.

The *span* or *closure of* a subset $A$ of $M$ is the minimal span in $M$ which contains $A$. Clearly, the span of $A$, which we always denote by $S(A)$, is unique. Where $A$ is a subset of column vectors in a matrix $M$ of column vectors, $S(A)$ is all the columns in $M$ which are linear combinations of $A$.

The terms above are used extensively in section 1.5 and section 1.6 to prove theorem 1. The terms below, through (4) and (5), are used extensively in Part II.

A subset $A$ of $M$ is said *to span* a subset $K$ of $M$ when $K \subset S(A)$. It follows from proposition 4, to come, that $A$ spans $K$ in $M$ if and only if for each element $e \in K$ either $e \in A$ or their is a circuit $C$ of $M$ such that $e \in C$ and $C - e \subset A$.

(4) The family of spanning sets of $M$. A *spanning subset* of $M$ is a subset of $M$ which spans $M$—in other words, a subset of $M$ whose span is $M$.

(5) The family of bases of $M$. A *base* of $M$ is a maximal independent set of $M$. A base can also be defined as a minimal spanning set of $M$.

The terms in (1), (2), and (5) are taken from Whitney [8]. The terms "closed set" and "span of $A$" are taken from Lehman [3]. There is an alternative terminology due to Tutte [7]. Since these are major sources on matroids, it is worthwhile to set down the relationship. To do so it is necessary to invoke the much used notion of "dual matroid," though it is not used here or in Part II. Papers [3], [7], and [8] show that the set-complements of the bases in a matroid $M$ are the bases of a so-called *dual matroid* $M^*$.

The bases of matroid $M$ are called by Tutte the *dendroids* of $M$. The elements of $M$ are called by Tutte the *cells* of $M$. The independent sets of $M$ are called by Lehman the *trees* of $M$.

The circuits of a matroid $M$ are what Tutte calls the *atoms* of dual matroid $M^*$. The circuits of $M^*$ are the atoms of $M$. Thus here is another special family of subsets of a matroid $M$.

(6) The family of atoms (dual circuits) in $M$.

The rows of a matrix $N_0$, under addition and subtraction, generate a group of row vectors which Tutte calls a *chain-group*, say the *chain-group* $N$ of matrix $N_0$. The matroid $M$ of matrix $N_0$ is of course an invarient of chain-group $N$, and it is what Tutte calls the matroid of chain-group $N$. An *atom* of $M$ of $N$ is defined as a set of elements in $M$ which corresponds to a minimal nonempty set of row-vector components

such that there is some member of chain-group $N$ which has its nonzero values in precisely these components. The row-vectors orthogonal to each row of matrix $N_0$ form another chain-group, say $N^*$. Its matroid is $M^*$, the dual of $M$. Atoms of $M^*$ by definition correspond to minimal dependent sets of columns in matrix $N_0$. That is, they are the circuits of the matroid $M$ of $N_0$.

Tutte defines a flat of matroid $M$ to be a union of atoms of $M$, or the empty set. It can be shown that a flat of $M$ is the set-complement of a span (closed set) in $M$, and conversely.

Where $A$ is a subset of elements in $M$, Tutte denotes by $M \cdot A$ what here is called the submatroid $A$ of $M$ (following Whitney). The meanings of the rank $r(M)$ of matroid $M$ coincide, and Tutte denotes by $r(M \cdot A)$ what here is called the rank $r(A)$ of set $A$ in $M$ (following Whitney). However, for a set $A$, what Tutte denotes by *r A is not* $r(A) = r(M \cdot A)$ but "$r(M \times A)$" which is used in Part II.

### 1.5. The Lemmas

In the proof of theorem 1 we will use axiom 1 and the following axiom 2′ for matroids instead of axioms 1 and 2.

AXIOM 2′: *The union of any independent set and any element contains at most one circuit (minimal dependent set).*

PROPOSITION 1: *Axioms 1 and 2′ are equivalent to axioms 1 and 2.*

PROOF: Assuming 1 and 2, suppose independent set $I$ together with element $e$ contains two distinct circuits $C_1$ and $C_2$. Assume $I$ is minimal for this possibility. $e \epsilon C_1 \cap C_2$. There is an element $e_1 \epsilon C_1 - C_2$ and an element $e_2 \epsilon C_2 - C_1$. Set $I \cup e - e_1 - e_2$ is independent since otherwise $(I - e_1)$ is a smaller independent set than $I$ for which $(I - e_1) \cup e$ contains more than one circuit. Set $I$ and set $I \cup e - e_1 - e_2$ are maximal independent subsets of set $I \cup e$. This contradicts axiom 2.

Assuming 1 and 2′, suppose $I_1$ and $I_2$ are both maximal independent subsets of a set $A$ such that $|I_1| < |I_2|$. Assume $I_1 \cup I_2$ is minimal for this possibility. There is an $e_1$ in $I_1 - I_2$ and $I_2 \cup e_1$ is dependent. By 2′, $I_2 \cup e_1$ contains a unique circuit $C$ which must contain some element $e_2$ not in $I_1$. Since $I_2$ is larger than $I_1$ it must contain another element besides $e_2$ not in $I_1$ and hence $I_2 \cup I_1 - e_2$ is dependent. Therefore, since $I_2 \cup e_1 - e_2$ is independent, there is some $I_1'$ such that $e_1 \epsilon I_1' \subset I_1 - I_2$ and such that $I_2' = I_2 \cup I_1' - e_2$ is maximal independent in $A$. Because $I_1'$ contains an element not in $I_2$, $|I_2'| \geq |I_2| > |I_1|$. However, since $I_1 \cup I_2'$ is a proper subset of $I_1 \cup I_2$, this contradicts the minimality assumption for $I_1 \cup I_2$. The proposition is proved.

PROPOSITION 2: *Axioms 1 and 2′ are equivalent to the following axioms,* $1_c$ *and* $2_c$*, for a matroid in terms of its circuits (where starting with circuits, independent sets are defined as sets containing no circuits).*

AXIOM $1_c$: No circuit contains another circuit.

AXIOM $2_c$: If distinct circuits $C_1$ and $C_2$ both contain an element $e$ then $C_1 \cup C_2 - e$ contains a circuit.

A proof of proposition 2 is obvious.

The next very useful proposition is taken in [7] and [8] to be an axiom instead of $2_c$. Alfred Lehman discovered that $1_c$ and $2_c$ suffice.

PROPOSITION 3: If $C_1$ and $C_2$ are circuits of a matroid $M$ with an element $e \epsilon C_1 \cap C_2$ and an element $a \epsilon C_1 - C_2$, then there is a circuit $C$ such that

$$a \epsilon C \subset C_1 \cup C_2 - e.$$

PROOF (Lehman): Assuming $1_c$ and $2_c$, suppose $C_1$, $C_2$, $a$, and $e$ are such that the theorem is false and $C_1 \cup C_2$ is minimal. There is a circuit $C_3 \subset C_1 \cup C_2 - e$, but $a \notin C_3$. There is an element $b \epsilon C_3 \cap (C_2 - C_1)$. By minimality of $C_1 \cup C_2$ for falsity of the theorem and since $a \notin C_2 \cup C_3$, there is a circuit $C_4$ such that $e \epsilon C_4 \subset C_2 \cup C_3 - b$. Again by the minimality and since $b \notin C_1 \cup C_4$, there is a circuit $C$ such that

$$a \epsilon C \subset C_1 \cup C_4 - e \subset C_1 \cup C_2 - e,$$

contradicting the falsity of the theorem.

PROPOSITION 4: *An element* e *of a matroid* M *is in the span* S(A) *of a set* A *in* M *if and only if* e *is in* A *or there is a circuit* C *of* M *for which* C − A = e.

PROOF: The "if" part of the theorem is asserted in the definition of span. Assuming the "only if" part false, by the definition of span there must be an $A$ and $e \epsilon S(A) - A$ for which there is no $C$ with $C - A = e$ but for which there is a $C$ and nonempty $E$ with $C - (A \cup E) = e$ where for each $e' \epsilon E$ there is a $C'$ with $C' - A = e'$. Assume $E$ to be minimal so that $E \subset C$. By prop. 3, for any $e'$ and $C'$ there is a $C_1$ for which $e \epsilon C_1 \subset C \cup C' - e'$. Hence, $C_1 - (A \cup E_1) = e$ where $E_1$ is a proper subset of $E$, contradicting the minimality of $E$.

Besides axioms 1 and 2′ and the definitions of circuit and span, the only other fact on matroids used to prove theorem 1 is

PROPOSITION 5: *The span of a set* A *in a matroid* M *is the (unique) maximal set* S *in* M *which contains* A *and which has the same rank as* A.

In particular the additional fact used in proving theorem 1 is that the span of an independent set $I$ has rank equal to the cardinality of $I$.

PROOF OF PROP 5: If, for $S(A)$ the span of $A$, $r(S(A)) > r(A)$, then by axiom 2 a base $I$ of A is not a base of S(A), i.e., there is an element $e \epsilon S(A) - I$ such that $I \cup e$ is independent. By prop. 4, $e$ is not in the span $S(I)$ of $I$ but $A$ is in $S(I)$. Since the span of a set is the minimal span containing the set, $S(A) \subset S(I)$. Thus, by contradiction, $r(S(A)) = r(A)$.

Let $e \epsilon S'(A)$ where $A \subset S'(A)$ and $r(A) = r(S'(A))$. Then, where $I$ is a base of $A$, either $e \epsilon I$ or $e \cup I$ is dependent. Thus $e \epsilon S(A)$. Therefore, $S(A)$ is the unique maximal set where $A \subset S(A)$ and $r(S(A)) = r(A)$.

### 1.6. The Main Proof

PROOF of theorem 1 (the "if" part): Assume that for every subset $A$ of matroid $M$, $|A| \leq k \cdot r(A)$. Actually, it is sufficient that for every span $S$ in $M$, $|S| \leq k \cdot r(S)$.

The goal is to get all the elements of $M$ into just $k$ independent sets of $M$. Let $F$ be a family of $k$ mutually disjoint independent sets of $M$. Any number of these sets may be empty. These sets are to be regarded as labeled so that each may be altered in the course of the proof while still maintaining its label-identity. Suppose there is an element $x$ of $M$ such that $\cup\{I: I\epsilon F\} \subset M-x$. We shall see how to rearrange elements among the members of $F$ to make room for $x$ in one of them while preserving the independence (and mutual disjointness) of them all. The process can be repeated until each element of $M$ is in a member of $F$. Thus the theorem will be proved.

Implementing this proof to an algorithm for partitioning (if possible) a matroid $M$ into $k$ independent sets is quite straight-forward as long as an algorithm is known for the following: for any $A\subset M$ and $e\epsilon M$, find a circuit $C$ such that $e\epsilon C\subset A\cup e$ or else determine that there is none. In the algorithm for partitioning $M$, one of course would *not* first verify $|A| \leq k\cdot r(A)$ for all $A\subseteq M$, but would simply proceed on the assumption that it is true and then stop if a contradiction arises.

If every member of $F$ contained as many as $r(M)$ elements, then since they are disjoint and do not contain $x$, the union of all $k$ of them together with $x$, which is a subset of $M$, would have cardinality greater than $k\cdot r(M)$. However, $|M|\leq k\cdot r(M)$. Hence there is an $I_1\epsilon F$ for which $|I_1| < r(M)$. Similarly, $x\epsilon S_1 = S(I_1)$ implies that there is an $I_2\epsilon F$ for which $|I_2\cap S_1| < r(S_1)$, since if each member of $F$ had $r(S_1)$ elements in $S_1$, then their union together with $x$ would be more than $k\cdot r(S_1)$ elements in $S_1$, but $|S_1| \leq k\cdot r(S_1)$.

Denoting $M$ by $S_0$, then likewise in general

$$x\epsilon S_i = S(I_i\cap S_{i-1})$$

implies that there is an $I_{i+1}\epsilon F$ for which $|I_{i+1}\cap S_i| < r(S_i)$, since $|S_i|\leq k\cdot r(S_i)$. These $I_i$'s are not necessarily distinct members of $F$.

Where

$$S_{i+1} = S(I_{i+1}\cap S_i),$$

we have

$$r(S_{i+1}) < r(S_i).$$

Since rank is a nonnegative integer, we must eventually reach an integer $h$ for which

$$x\notin S_h = S(I_h\cap S_{h-1})$$

and

$$x\epsilon S_i \text{ for } i = 1, \ldots, h-1.$$

By construction, $S_1\supset S_2\supset \ldots \supset S_h$.

If $I_h\cup x$ is independent then replacing $I_h$ by $I_h\cup x$ disposes of $x$. Otherwise there is a unique circuit $C\subset I_h\cup x$. Since $C-x\subset S_{h-1}$ would imply $x\epsilon S_h = S(I_h\cap S_{h-1})$, there is an $x_1\epsilon C-x$ such that $x_1\notin S_{h-1}$. We replace $I_h$ in $F$ by independent set $I_h\cup x-x_1$. The new family is still called $F$ and the new set carries the label-identity in $F$ which $I_h$ had. This and the following informal conventions are used simply to avoid introducing a lot more indices. Any other $I_i$ which was the same member of $F$ as $I_h$ is now $I_h\cup x-x_1$. We will distinguish between the current $I_i$ and the original $I_i$. The $S_i$'s do not change.

We have disposed of $x$ and now we must find a place for $x_1$ in some member of $F$. Since $x_1\notin S_{h-1}$ and $x_1\epsilon S_0$, and since the $S_i$'s are monotonically nested, there is some index $i(1)\leq h-1$ for which

$$x_1\notin S_{i(1)} \text{ and } x_1\epsilon S_{i(1)-1}.$$

Denote $h$ by $i(0)$ and denote $x$ by $x_0$. Assume inductively that $x_0\notin S_{i(0)}$, $x_0\epsilon S_{i(0)-1}$, $x_1\notin S_{i(1)}$, $x_1\epsilon S_{i(1)-1}$, . . ., $x_j\notin S_{i(j)}$, $x_j\epsilon S_{i(j)-1}$, where $i(0) > i(1) > \ldots > i(j)$. Assume further that $I_{i(0)}$ was replaced in $F$ by $I_{i(0)}\cup x_0 - x_1$, then $I_{i(1)}$ was replaced in $F$ by $I_{i(1)}\cup x_1 - x_2$, . . ., and then $I_{i(j-1)}$ was replaced in $F$ by $I_{i(j-1)}\cup x_{j-1} - x_j$; where $x_1\epsilon C_0\subset I_{i(0)}\cup x_0$, $x_2\epsilon C_1\subset I_{i(1)}\cup x_1$, . . ., and $x_j\epsilon C_{j-1}\subset I_{i(j-1)}\cup x_{j-1}$.

Suppose there is a circuit $C_j\subset I_{i(j)}\cup x_j$. Set $I_{i(j)}$ might have the same label-identity in $F$ as $I_{i(q)}$ for several values of $q<j$, and so the contents of $I_{i(j)}$ may have changed several times since the original $I_{i(j)}$ which gave rise to $S_{i(j)} = S(I_{i(j)}\cap S_{i(j)-1})$. In particular, $x_q$ for some $q<j$ may have been adjoined to $I_{i(j)}$. However, by the induction hypothesis any such $x_q$ is contained in $S_{i(q)-1}$ and thus in $S_{i(j)}$.

Therefore all elements of $C_j - x_j$ which are not in the original $I_{i(j)}$ are in $S_{i(j)}$. By definition of $S_{i(j)}$, all elements of the original $I_{i(j)}$ which are in $S_{i(j)-1}$ are also in $S_{i(j)}$. Thus if all elements of $C_j - x_j$ are in $S_{i(j)-1}$ then they are all in $S_{i(j)}$, but since $S_{i(j)}$ is a span then $x_j$ also would be in $S_{i(j)}$, contradicting the inductive hypothesis. Hence, there exists some element $x_{j+1}$ of $C_j$ such that $x_{j+1}\notin S_{i(j)-1}$. Since $x_{j+1}\epsilon S_0$, there is some $i(j+1) < i(j)$ such that $x_{j+1}\notin S_{i(j+1)}$ and $x_{j+1}\epsilon S_{i(j+1)-1}$.

Therefore when there exists a $C_j$, we repeat the inductive step by replacing $I_{i(j)}$ by $I_{i(j)}\cup x_j - x_{j+1}$.

Since $i(0) > i(1) > \ldots$, eventually we must reach an $i(j)$ for which there is no $C_j\subset I_{i(j)}\cup x_j$. Then we can replace $I_{i(j)}$ in $F$ by independent set $I_{i(j)}\cup x_j$ without having to displace another element $x_{j+1}$. End of proof.

### 1.7. Corollary

For the special case where $M$ is the matroid of a graph $G$, theorem 1 can be simplified somewhat:

COROLLARY (Nash-Williams [5]): *The edges of a graph* G *can be colored with as few as* k *colors so that no circuit of* G *is all one color, if and only if there is no subset* U *of nodes in* G *such that, where* $E_U$ *is the set of edges in* G *which have both ends in* U,

$$|E_U| > k(|U|-1).$$

Symbols $|U|$ and $|E_U|$ denote, respectively, the cardinalities of $U$ and $E_U$.

Not every subset $A$ of elements in the matroid $M(G)$ of $G$, nor even every closed set $A$ of elements in $M(G)$, corresponds to a set of edges of type $E_U$. However, the relation $|A| \leqq k \cdot r(A)$ for every set $A$ corresponding to a set $E_U$ of edges which form a connected subgraph of $G$ implies the relation for every subset $A$ of elements in $M(G)$.

The corollary follows (we omit the proof) from theorem 1 by using the following characterization of the rank function of a graph due to Whitney:

The rank $r(E)$ of any subset $E$ of edges in $G$, i.e., the rank of the matroid subset corresponding to $E$, equals the number of nodes minus the number of connected components in the subgraph, $G \cdot E$, consisting of the edges $E$ and the nodes they meet, or equivalently in the subgraph, $G : E$, consisting of the edges $E$ and all the nodes of $G$. The notation $G \cdot E$ and $G : E$ is due to Tutte, chapter III of [7].

## References


[1] G. Birkhoff, Abstract linear dependence and lattices, Amer. J. Math. 57, 800–804 (1935).

[2] H. Crapo, Single-element Extensions of Matroids, J. Res. NBS 69B (Math. and Math. Phys.) No. 1,

[3] A. Lehman, A solution of the Shannon switching game, Univ. of Wisc. Math. Research Center Report #308, 1962, to appear in the SIAM Journal.

[4] C. St. J. A. Nash-Williams, Edge-disjoint spanning trees of finite graphs, J. London Math. Soc. **36,** 445–450 (1961).

[5] C. St. J. A. Nash-Williams, Decomposition of finite graphs into forests, J. London Math. Soc. **39,** 12 (1964).

[6] W. T. Tutte, On the problem of decomposing a graph into $n$ connected factors, J. London Math. Soc. **36,** 221–230 (1961).

[7] W. T. Tutte, Lectures on matroids, J. Res. NBS 69B (Math. and Math. Phys.) No. 1,

[8] H. Whitney, On the abstract properties of linear dependence, Amer. J. Math. 57, 509–533 (1935).


(Paper 69B1–134)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Lehman's Switching Game and a Theorem of Tutte and Nash-Williams[1]

Jack Edmonds

(December 1, 1964)

The results cited in the title are unified by the following theorem: For any matroid $M$ and any subsets $N$ and $K$ of elements in $M$, there exist as many as $k$ disjoint subsets of $N$ which span $K$ and which span each other if and only if there is no contraction matroid $M \times A$ where $N \cap A$ partitions into as few as $k$ sets such that each is independent in $M \times A$ and such that at least one of them does not span $K \cap A$ in $M \times A$.

## 2.1. The Problem

A. Lehman [3][2] posed the following game to be played between two players on any given matroid $M$ with a distinguished element $e$. The players are called the cut player and the short player. They take turns and (to be explicit) the cut player goes first. Each player in his turn tags an element of $M$, other than $e$, not already tagged. The short player wins if he tags a set of elements which span $e$. The cut player wins otherwise—that is, the cut player wins if the elements, other than $e$, which he has not tagged do not span $e$.

The game, determined by $M$ and $e$, is called a *short game* if the short player can win against any strategy of the cut player. We will call the game *nonshort* if the cut player can win against any strategy of the short player. Clearly a game is one or the other. For any $M$ and $e$, Lehman characterizes short games and describes a winning strategy for the short player.

Recall from section 1.4 that a set $T$ in a matroid $M$ is said to span a set $A$ in $M$ if for every $e \in A$, either $e \in T$ or there is a circuit $C$ of $M$ such that $C - e \subset T$. Recall that a base $B$ of $M$ is a set which spans $M$ (e.g., a spanning set of $M$) and which also is independent.

Where the game is played on a graph $G$, it is not necessary to have an edge corresponding to $e$ but sufficient to have two distinguished "terminal" nodes, $v_1$ and $v_2$, which would be the ends of $e$. Here, the goal of the short player is to tag a set of edges which contains a path of edges joining $v_1$ to $v_2$. The goal of the cut player is to tag a set of edges which separates $v_1$ from $v_2$.

A theorem due independently to Tutte [6] and Nash-Williams [4] characterizes for any graph $G$ the maximum number of edge-wise disjoint subgraphs, each connected and containing all nodes of $G$, into which the edges of $G$ can be partitioned. For a connected graph $G$, the edges of a connected subgraph which contains every node of $G$ correspond to the elements of a spanning set of the matroid of $G$, and conversely.

The purpose of the present note is to unify these two theories. Theorem 2 states the straightforward generalization to matroids of the Tutte and Nash-Williams theorem. Theorem 3 is Lehman's main theorem characterizing short games. Theorem 4 is an analogous theorem characterizing nonshort games. (Lehman characterizes nonshort games indirectly by using "dual matroids" which we avoid.) Theorem 5, for the case where $K = N = M$, yields theorem 2. For the case where $k = 2$, it yields the "only if" parts of theorems 3 and 4. The "if" parts of theorems 3 and 4 are proved by describing the winning strategies when the respective conditions hold (in one case this follows Lehman, [3]).

Theorem 1 in section 1.3 and theorem 2 are in a sense dual to each other but not in the usual matroid sense. Each can be proved from the other. We use theorem 1 to prove theorem 5.

Theorem 5 appears interesting in itself. We call it the "cospanning-set theorem" after a main idea of Lehman's theory. For a graph $G$ with a prescribed subset of nodes called terminals, it gives a "good" characterization for the nonexistence of $k$ edge-wise disjoint connected subgraphs (e.g., trees), all with precisely the same set of nodes which includes the terminals.

If the matroid $M$ of the cospanning-set theorem is a finite set of vectors in a space $L$, then for given subsets $N$ and $K$ of $M$, the theorem provides a "good" characterization for the nonexistence of as many as $k$ disjoint subsets $N_i$ of $N$ and a subspace $L'$ of $L$ such that each $N_i$ exactly spans $L'$ and such that $L'$ contains $K$.

## 2.2. Contractions

We use the following inportant concept on matroids due to Tutte (ch. II of [7]). For any set $A$ of elements in a matroid $M$, define the circuits of $M \times A$ to be the minimal nonempty intersections of $A$ with circuits of $M$.

PROPOSITION 6: *The set of elements* A *and the circuits of* M × A *are a matroid (denoted by* M × A), *called the contraction of* M *to* A.

PROOF: Axioms $1_c$ and $2_c$ for $M \times A$ follow immediately from prop. 3 for $M$.

[1] This paper is a sequel to the preceding one, "Minimum Partition of a Matroid Into Independent Subsets." The numbering system there, including references, is continued here. This work was supported by the Army Research Office (Durham) and the Defense Communications Agency.

[2] Figures in brackets indicate the literature references on page 72.

COROLLARY: *Where* A *and* $\overline{\text{A}}$ *are complementary subsets of matroid* M, $\overline{\text{A}}$ *is closed (a span) in* M *if and only if matroid* M × A *contains no "loops," that is elements of rank zero.*

PROPOSITION 7: *Where* K *and* A *are subsets of matroid* M, *subset* T′ *of* A *spans* K∩A *in* M × A *if and only if there is a subset* T *of* M *such that* T′ = T∩A *and such that* T *spans* K *in* M.

COROLLARY: *The spanning sets of matroid* M × A *are precisely the intersections of* A *with spanning sets of* M.

PROOF OF PROP. 7: Suppose $T' = T \cap A$ where $T$ spans $K$ in $M$. Since $T$ spans $K$, for any element $e$ in $K \cap A$, either $e \epsilon T$ or there is a circuit $C$ in $M$ such that $C - e \subset T$. If $e \epsilon T$ then $e \epsilon T'$ and hence $T'$ spans $e$ in $M \times A$. If there is a $C$ then, by definition of $M \times A$, there is a circuit $C'$ of $M \times A$ such that $e \epsilon C' \subset C$. It follows that $C' - e \subset T'$ and hence $T'$ spans $e$ in $M \times A$. Thus, the "if" part is proved.

Suppose subset $T'$ of $A$ spans $K \cap A$ in $M \times A$. Let $T = T' \cup \overline{A}$ where $\overline{A}$ is the complement of $A$ in $M$. Then $T' = T \cap A$. Let $e$ be any element of $K$. If $e \epsilon T$, then $T$ spans $e$ in $M$. Otherwise, $e \epsilon K \cap A$, *and* $e \notin T'$. Since $T'$ spans $e$ in $M \times A$, there is a circuit $C'$ of $M \times A$ such that $e \epsilon C'$ and $C' - e \subset T'$. By definition of $M \times A$, there is a circuit $C$ of $M$ such that $C' = C \cap A$. Therefore, $T$ spans $e$, since $e \epsilon C$ and $C - e \subset T$. Thus, the "only if" part is proved.

Tutte uses $M \cdot A$ to denote what we mean by the submatroid $A$ of $M$; he does not follow Whitney's informality of letting $A$ mean both a matroid and its set of elements. We will use Tutte's notation and also, where convenient, we will depart from it again by referring to $M \times A$ simply as the contraction matroid, $A$, of $M$ just as we refer to $M \cdot A$ as the submatroid, $A$, of $M$. Also, $A$ denotes the elements of either.

Where $M(G)$ is the matroid of graph $G$, the matroid of a subgraph $H$ of a graph $G$ is the submatroid of $M(G)$ which contains the elements corresponding to the edges of $H$, and conversely. The matroid of a "contraction graph" $H$ of $G$ is the contraction of $M(G)$ which contains the elements corresponding to the edges of $H$, and conversely.

The most instructive way to describe the meaning of contraction graph is visually. The contraction graph $H$ of $G$ whose edges are the set $H$ of edges in $G$ is the graph obtained from $G$ by contracting to a point each edge of $G$ not in $H$.

It should be pointed out that in order for there to be a contraction $H$ of $G$ for every subset $H$ of edges in $G$, we must extend our meaning of graph (in sec. 1.1) to graphs which include edges which "meet the same node at both ends." These "loop" edges are circuits by themselves; they correspond to matroid elements which are not contained in any independent set of the matroid. This sort of matroid element corresponds in a matrix to a column of all zeros. In a matrix $N(G)$, a loop of graph $G$ can be represented by a column of $N(G)$ which contains a 2 in the row corresponding to the node met and which contains zeros elsewhere. Relative to the matroid structure, the column is all zeros, mod. 2.

We have pointed out how any contraction of the matroid of a graph can be represented as the matroid of a graph. It is also possible to represent any contraction of the matroid of a matrix as the matroid of a matrix.

By *deleting* (or *cutting*) set of elements $A$ in matroid $M$, we mean replacing matroid $M$ by its submatroid on the set $M - A$. By *contracting* (or *shorting*) set of elements $A$ in matroid $M$, we mean replacing matroid $M$ by its contraction to the set $M - A$. Clearly, from the definition of submatroid, we can get a submatroid $M'$ of $M$ by deleting the elements of $M$ not in $M'$ one after another in any order. Clearly, from the corollary to prop. 7, we can get a contraction matroid $M'$ of $M$ by contracting the elements of $M$ not in $M'$ one after another in any order. It can be proved that for any elements $a$ and $b$ in a matroid $M$, deleting $a$ and then contracting $b$ is the same as contracting $b$ and then deleting $a$. The proof is omitted. These results can be summarized by the following:

PROPOSITION 8: *The operations of deleting certain elements together with the operations of contracting certain other elements in a matroid are associative and commutative.*

The above proposition is equivalent to Tutte's identities 3.33 in [7]. Tutte defines a *minor* of a matroid $M$ to be any matroid obtained from $M$ by deleting certain elements and contracting certain other elements in $M$.

The following theorem is presented by Tutte (theorem 3.53 of [7]) in terms of "dendroids."

PROPOSITION 9: *If* A *and* $\overline{\text{A}}$ *are complementary sets of elements in matroid* M, *then the elements in a base of* M × A *together with the elements in a base of* $\text{M} \cdot \overline{\text{A}}$ *are the elements in a base of* M.

Proof omitted.

COROLLARY: $r(\text{M} \cdot \text{A}) + r(\text{M} \times \overline{\text{A}}) = r(\text{M})$.

We have been calling $r(M \cdot A)$ the rank $r(A)$ of set $A$ in matroid $M$. We denote $r(M \times A)$ as function $t(A)$ of sets $A$ in matroid $M$.

The following theorem, which for the case of connected graphs is the one due to Tutte and Nash-Williams, completely parallels theorem 1. The "if" part of theorem 2 follows immediately from the "if" part of theorem 5 (where $M = N = K$).

THEOREM 2: *The elements of a matroid* M *can be partitioned into as many as* k *sets, each a spanning set of* M, *if and only if there is no subset* A *of elements of* M *for which*

$$|\text{A}| < \text{k} \cdot \text{t(A)}.$$

Any contraction graph of a connected graph is connected. Using the last paragraph of 1.7, observe that where $M$ is the matroid of a connected graph $G$, $t(A)$ is the number of nodes minus one of a contraction graph of $G$, and $|A|$ is the number of edges in that contraction graph.

Notice that, since $t(A) = r(M) - r(\overline{A})$, theorem 2 is easily stated without the notion of contraction.

To prove the "only if" part of theorem 2, assume that $M$ partitions into $k$ sets, each spanning $M$. By taking a subset of each of them, we get disjoint bases

$B_i (i = 1, \ldots, k)$. Let $A$ be any subset of $M$ and let $\bar{A}$ be its complement. Since $B_i$ is independent, $r(\bar{A}) \geq |\bar{A} \cap B_i|$. Since $B_i$ is a base, $|A \cap B_i| + |\bar{A} \cap B_i| = |B_i| = r(M)$. Combining the two gives $|A \cap B_i| \geq r(M) - r(\bar{A}) = t(A)$. Therefore $|A| \geq \Sigma_i |A \cap B_i| \geq k \cdot t(A)$.

### 2.3. Short Games

It turns out to be just as easy to analyze games where, for the graph case, any subset of nodes of $G$ are distinguished as terminals and the goal of the short player is to tag a set of edges in $G$ which contains the edges of a connected subgraph containing all the terminals. To interpret this game in matroid terms, adjoin to $G$ a set of new edges which form a connected graph $K$ containing precisely the terminals as nodes. Then relative to the matroid of graph $G \cup K$, the goal of the short player is to tag a set of elements corresponding to edges in $G$ which spans the set of elements corresponding to edges in $K$.

For any matroid $M$ and nonempty subsets $N$ and $K$, consider the game $L(M, N, K)$ where, as before, the cut player and short player take turns tagging different elements of $N$, the cut player going first. The short player wins if he tags a set of elements which span $K$. Otherwise, the cut player wins. Call $L(M, N, K)$ a short game if the short player can win against any strategy of the cut player.

Lehman's main theorem (explicitly for the case where $K$ is a single element) is

THEOREM 3: *$L(M, N, K)$ is a short game if and only if* N *contains two disjoint sets,* $A_0$ *and* $B_0$, *of elements which span each other and which span* K.

Notice that in the two-terminal graph case, the short player wants to get a path joining the terminals. The structure characterizing when he can is two edgewise disjoint trees each containing the terminals and each containing *precisely* the same nodes as the other.

Lehman calls two (or more) sets which span each other *cospanning*. Let us verify that two disjoint cospanning sets $A_0$ and $B_0$ in $N$ which span $K$ provide a winning strategy in the game $L(M, N, K)$ for the short player. All that we need consider is the span $M_0 = S(A_0) = S(B_0)$ in $M$. Clearly, we can take $A_0$ and $B_0$ to be bases of submatroid $M_0$; assume that they are. If the cut player tags an element not in $A_0 \cup B_0$, we can pretend that at the same time he also tags some element of $A_0 \cup B_0$. Clearly, the short player would not be taking an illegal advantage by pretending this. Therefore, suppose the cut player in his first turn tags element $a_0$ in $A_0$.

By axiom 2 (in 1.1) there is an element $b_0$ of $B_0$ such that $(A_0 - a_0) \cup b_0$ is a base of $M_0$. The short player should tag an element $b_0$. It follows from prop. 7 that disjoint sets $A_1 = A_0 - a_0$ and $B_1 = B_0 - b_0$ are spanning sets of the contraction matroid $M_1 = M_0 - b_0$ of $M_0$.

Since it is the cut player's turn again, the situation of $A_1$ and $B_1$ relative to $M_1$ is as it was for $A_0$ and $B_0$ relative to $M_0$ except that $M_1$ is smaller. Assuming there is a strategy for the succeeding turns whereby the short player can tag a set of elements which contains a base $T$ of reduced matroid $M_1$, then by prop. 9 the set $T \cup b_0$ of elements, which the short player will have tagged, is a base of matroid $M_0$ and hence spans set $K$ in matroid $M$.

When $B_0$ contains only one element $b_0$, then $b_0$ itself spans $M_0$ and $K$. Hence, by induction on the number of elements, we have a winning strategy for the short player. This proves the "if" part of theorem 3. The harder "only if" part will follow from theorem 4 and theorem 5.

### 2.4. Nonshort Games

The notion of contraction can always be used in place of the more familiar notion of "matroid duality," and conversely, because of a theorem (3.27 of [7]) relating the contraction matroids of an $M$ to the submatroids of the "dual to $M$." Sometimes one notion is convenient, sometimes the other. We do not use duality here. Lehman in treating the same topic uses mainly duality.

Lehman's interpretation of his dual results characterizing when the cut player can win for the case of graphs does not directly provide a "good" characterization in the sense of the absolute supervisor. Clearly his characterization of a short game is good in the case of graphs. However, he does not give the following analogous characterization for nonshort games. (Compare Lehman's theorem (26) and its graph interpretation with our theorem 4 and its contraction graph interpretation. See also the comment on his theorem (26) which follows his theorem (29).)

THEOREM 4: L(M, N, K) *is a non-short game if and only if there is a contraction matroid* M′ *of matroid* M *where set* $N' = N \cap M'$ *can be partitioned into two sets* $I_1$ *and* $I_2$ *such that* $I_1$ *and* $I_2$ *are both independent in* M′ *and such that* $I_2$ *does not span the set* $K' = K \cap M'$ *in* M′.

Let us verify that an $M'$, $I_1$, and $I_2$ provide a winning strategy for the cut player in game $L(M, N, K)$. If $I_1$ does not span $K'$ in $M'$ then the cut player can tag anything on his first turn. Otherwise, he should tag an element $e_1$ in $I_1$ such that $I_1 - e_1$ does not span $K'$ in $M'$.

Since $I_2$ does not span $K'$, there is an element $e \in K'$ such that $r(e) \neq 0$. If $e \in I_1$, then $e$ is an element $e_1$. Otherwise, by axiom 2′ there is a *unique* circuit $C$ in $I_1 \cup e$ and so any element of $C - e$ is an element $e_1$. ($C - e$ is not empty since $r(e) \neq 0$.) Now neither the untagged elements $I_1' = I_1 - e_1$ of $I_1$ nor the untagged elements $I_2' = I_2$ of $I_2$ span $K'$ in contraction matroid $M'$.

Even if the short player tags an element not in $M'$, clearly the cut player is not taking an illegal advantage by pretending the short player also tags an untagged element in $M'$ if there are any. Therefore, assume the short player does tag one, say $e_2$ in $I_2'$. Consider the contraction matroid $M'' = M' - e_2$ of $M$. (By prop. 8, the contraction of matroid $M'$ to set $M''$ is the same as the contraction of matroid $M$ to set $M''$.) By the circuit definition of contraction matroid, set $I_2'' = I_2' - e_2$ will be independent in $M''$ and will not span $K'' = K' \cap M''$ in $M''$.

Again by the definition of contraction matroid, if $e_2$ is not in the span of $I_1'$ in $M'$ then $I_1'$ is independent in $M''$. In this case, the cut player should tag some ele-

ment $e_1'$ such that $I_1''=I_1'-e_1'$ does not span $K''$ in $M''$.

By the definition of contraction matroid and prop. 3, if $e_2$ is in the span of $I_1'$ in matroid $M'$ then set $I_1'$ contains just one circuit of contraction $M''$ and does not span $K''$ in $M''$. In this case, the cut player should tag some element $e_1'$ in the one circuit of $I_1'$ in $M''$, so that $I_1''=I_1'-e_1'$ is independent in $M''$.

Thus in either case, after the cut player takes his second turn, the untagged elements of $M''$ partition into sets $I_1''$ and $I_2''$ where, in contraction $M''$, both are independent and neither spans $K''$. There are no elements tagged by the short player in $M''$. The situation is identical to the one in $M'$ right after the cut player took his first turn except that $M''$ has fewer untagged elements.

Hence, by induction on the number of untagged elements in the contraction matroid, if the cut player tags as described, he eventually reaches a contraction matroid $M^{(h)}$ in which all the elements are tagged by him, and yet for which there is an $e\epsilon K^{(h)}=M^{(h)}\cap K$ such that $r(e)\neq 0$ in $M^{(h)}$ (since $K^{(h)}$ is not spanned by the empty $I_1^{(h)}$ or the empty $I_2^{(h)}$). The cut player will then have won the game, because for the short player to win he must tag a set, say $T$, which spans $K$ in matroid $M$. By prop. 7, for any such $T$ and any set $M^{(h)}$ in $M$, $T\cap M^{(h)}$ must span $K^{(h)}=K\cap M^{(h)}$ in matroid $M^{(h)}$, which is impossible. This proves the "if" part of theorem 4.

## 2.5. Cospanning-Sets Theorem

We still have to prove the "only if" parts of theorems 3 and 4. They follow immediately from theorem 5 (for the case $k=2$). We proved the part of theorem 3 which says "$P\Rightarrow$ ($L$ is a short game)". We proved the part of theorem 4 which says "$Q\Rightarrow$ ($L$ is not a short game)". Theorem 5 says "$P\Leftrightarrow$ not $Q$". Logic yields that "($L$ is a short game)$\Rightarrow P$" and "($L$ is not a short game)$\Rightarrow Q$".

THEOREM 5: *For any matroid* M *and any subsets* N *and* K *of elements in* M, *there exist as many as* k *disjoint subsets of* N *which span each other and which span* K, *if any only if there is no contraction matroid* M′ *of* M *where* N∩M′ *partitions into as few as* k *sets such that each is independent in* M′ *and such that at least one of them does not span* K∩M′

PROOF: The "only if" part of theorem 5 follows from the "if" part of prop. 7. Suppose in matroid $M$ there exist $k$ disjoint subsets $T_i$ of $N\subset M$, which span each other and which span $K\subset M$. Let $M'$ be any contraction of $M$. Where a set $T_i$ is the $T$ of prop. 7; where set $M'$ is the $A$ and matroid $M'$ is the $M\times A$ of prop. 7; and where $S$, the span (closure) in $M$ of each $T_i$, is the $K$ of prop. 7; prop. 7 says that $T_i'=T_i\cap M'$ spans $S'=S\cap M'$ in matroid $M'$. Since each $T_i'$ spans $S'$, each $T_i'$ contains at least $r(S')$ elements where $r(S')$ is the rank of set $S'$ in matroid $M'$. Since all the sets $T_i'$ are mutually disjoint, $N\cap S'$ contains at least $k\cdot r(S')$ elements.

On the other hand, suppose $N\cap M'$ partitions into as few as $k$ independent sets $I_i$ of $M'$ where one of them $I_1$ does not span $K\cap M'$ and hence does not span $S'$. Since each $I_i'=I_i\cap S'$ is independent, each $I_i'$ contains at most $r(S')$ elements. Since $I_1'$ does not span $S'$, it contains fewer than $r(S')$ elements. Therefore $N\cap S'=\cup I_i'$ contains fewer than $k\cdot r(S')$ elements. Thus, the "only if" part of theorem 5 is proved.

The "if" part of theorem 5 follows from propositions 8 and 9 and theorem 1. Let $M$ be any matroid, let $N$ and $K$ be any subsets of $M$, and let $k$ be any positive integer. Suppose $A_0$ is a maximal subset of $N$ such that $|A_0|=k\cdot r(A_0)$ and $|A|\leq k\cdot r(A)$ for all $A\subset A_0$. Set $A_0$ may be empty. By theorem 1, $A_0$ partitions into $k$ independent sets, $I_i$. Since $|A_0|=k\cdot r(A_0)$, each $I_i$ must be a base of submatroid $A_0$ and of course also a base of $S(A_0)$, the span of $A_0$ in $M$.

Let $M'$ be the contraction matroid of $M$ obtained by contracting $S(A_0)$ in $M$. Suppose $A_1$ is a subset of $N'=N\cap M'$ such that $|A_1|=k\cdot r(A_1)$ and $|A|\leq k\cdot r(A)$ in matroid $M'$ for all $A\subset A_1$. Then like $A_0$ in $M$, $A_1$ partitions into $k$ bases $I_i'$ of submatroid $A_1$ of $M'$. By prop. 8, submatroid $A_1$ of $M'$ is the contraction to $A_1$ of the submatroid $A_1\cup S(A_0)$ of $M$. Call it minor $A_1$. By prop. 9, a base of minor $A_1$ together with a base of submatroid $S(A_0)$ of $M$ is a base of submatroid $A_1\cup S(A_0)$ of $M$.

In particular, by pairing the sets $I_i'$ one-to-one with the sets $I_i$, we get $k$ disjoint bases $I_i''=I_i'\cup I_i$ of submatroid $A_1\cup S(A_0)$ of $M$. Since $\cup I_i''=A_0\cup A_1\subset N$ and since the sets $I_i''$ span each other in $M$, $|A_0\cup A_1|=k\cdot r(A_0\cup A_1)$ and, by the "only if" part of theorem 1, $|A|\leq k\cdot r(A)$ in $M$ for all $A\subset A_0\cup A_1$. However, $A_0$ was taken to be maximal for this property, and hence $A_1$ is empty. Thus, matroid $M'$ contains no nonempty $A_1$, as defined.

Since $S(A_0)$ is closed in $M$, the matroid $M'$, obtained by contracting $S(A_0)$, contains no element of rank zero (corollary to prop. 6). Suppose $N'=N\cap M'$ contains a nonempty set $A_2$ such that $|A_2|\geq k\cdot r(A_2)$ in $M'$. Take $A_2$ to be minimal. By the nonexistence in $M'$ of a nonempty $A_1$ as described above, we have that $|A_2|>k\cdot r(A_2)$. Since there are no elements of rank zero, $A_2$ contains at least two elements. Deleting an element from $A_2$ to get a nonempty $A_3$, we have $|A_3|\geq k\cdot r(A_2)\geq k\cdot r(A_3)$ in $M'$, which contradicts the minimality of $A_2$. Therefore, for all nonempty subsets $A$ of $N'$, $|A|<k\cdot r(A)$ in $M'$.

Suppose some element $g\epsilon K$ is contained in matroid $M'$. Since $g$ does not have zero rank, there exists a matroid $M_h$, which contains the elements of $M'$ plus a new auxiliary element $h$, such that $h$ and $g$ form a circuit in $M_h$ and such that submatroid $M_h-h$ of $M_h$ is the matroid $M'$. It is easy to verify that $M_h$ is such a matroid where the circuits of $M_h$ are (1) the set consisting of $g$ and $h$, (2) the circuits of $M'$, and (3) sets $(C-g)\cup h$ where C is a circuit of $M'$ which contains $g$. Let $N_h=N'\cup h$. It follows from the relation $|A|<k\cdot r(A)$ in matroid $M'$ for all nonempty $A\subset N'$, that $|A_h|\leq k\cdot r(A_h)$ in $M_h$ for all $A_h$ in $N_h$.

Hence by theorem 1, $N_h$ can be partitioned into $k$ independent sets $I_i^h$ of $M_h$, including the set, say $I_1^h$, which contains $h$. In matroid $M'$ the set $I_1^h - h$ is independent and does not span $g$. All of the other sets $I_i^h$ are independent in $M'$. These sets $I_i^h$ and $I_1^h - h$ are a partition of $N'$.

Thus, if there is no such partition of $N' = N \cap M'$ for contraction $M'$ of $M$ then no element of $K$ is in $M'$. Thus $K \subset S(A_0)$. In this case, the $k$ bases $I_i$ of submatroid $S(A_0)$ of $M$ span each other and span $K$ in $M$. This completes the proof of theorem 5.

(Paper 69B1-135)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# On the Connection Between the Properties of Oriented Linear Graphs and Analyses of Lumped Physical Systems

Horace M. Trent[1]

U.S. Naval Research Laboratory

(February 17, 1965)

It is necessary at the outset of this lecture to apologize to this assemblage of mature theoreticians for talking about an essentially engineering subject. My excuse for doing so is that whereas my subject represents one of the most important current applications of the theory of linear graphs, it is all too clear from the literature that a number of misconceptions about it are being passed along from person to person. A word of warning is in order also. The subject I am to discuss is like a multifaced gem in that it has many facets, each of which can add to an appreciation of the beauty of the whole object. Indeed whole books have been written on a restricted aspect of the total subject; the application of linear graphs to electric networks for example. This being the case, it is foolish to think that I can give a definitive exposition of the subject in less than an hour. At most, it will be possible only to touch upon those ideas and concepts which either are rather basic or are often overlooked by some people who use the techniques.

It is well at the outset to bear in mind that the theory of linear graphs is used, in the application under discussion, as an aid, and as a unifying concept in the analysis of what can be called hereafter, an engineering system. In particular, the techniques are applicable to those engineering systems which can be described, with adequate precision, by a finite number of physical variables. This limitation assures us that we shall be dealing only with finite linear graphs as will become more evident later. These remarks suggest, or rather demand, that we look carefully at engineering analyses and extract from them those concepts and operations that are pertinent to the problem at hand. By this it is meant that our problem is to justify, in some logical fashion, just how the properties of linear graphs, which after all are only lines on a sheet of paper, can be used in a meaningful way in the analysis of a finite engineering system. Surely no one in this audience believes that a linear graph drawn on the blackboard in this room and, say, a motor-generator down the hall are the same object. And surely more than one person here is wondering why such a trite remark has been made. There are two reasons. In the first place some trite remarks emphasize fundamental concepts. Such a case is at hand for it is absolutely basic to understand that a linear graph and an engineering system under analysis are distinct objects. Second, the remark is made to counteract certain misconceptions that appear in the open literature. For example, one well known book contains a sentence which begins, "we define an electric network to be a linear graph, etc." For goodness sake, an electric network cannot be defined to be anything but an electric network. In fact, it couldn't care less how anyone defines it. That is another trite statement full of deep meaning for an electric network, or any engineering system for that matter, will go on performing its own function in its own peculiar way no matter what a theoretician down the hall has to say about it. Hence a theoretician, if he is a good one, will not try to warp the characteristics of a system to fit his own whims and notions; rather he can only hope that his whims and notions may have some reasonable correspondence to what goes on in the real system.

Using the example of the motor-generator set down the hall, it is obvious that anything a theoretician does at his desk has no reaction whatsoever with the system under analysis. In other words, the performance of the motor-generator is completely unaffected by anything an analyst does with a pencil and a pad of paper. These remarks point up the fact that an analysis of an engineering system is strictly a mental exercise indulged in by a theoretician. Failure to recognize this fact can lead to unrealistic concepts; it has done so not too infrequently in the past.

It should be apparent at this point that if a linear graph is to be useful at all in the analysis of an engineering system it is highly likely that it can do so only by virtue of an isomorphism between some, but not all, of its properties and some of the mental objects which are created in the course of analyzing a system. The remainder of the remarks to be made today show in outline how such an isomorphism can be established.

Let us begin by looking at the pertinent mental exercises indulged in by an analyst. The order in which these mental operations will be discussed is not necessarily the order in which they are done in a specific application.

First, some assumptions must be made about the nature of space and time. It is often assumed for

[1] Dr. Trent died December 16, 1964. He was Head of Applied Mathematics at NRL.



757-615 O-65-6

example that space is Euclidean and that time is isochronous. It would be good if it could be asserted that analysts are aware of their making such assumptions each time they do so; but unfortunately this is not the case. Nevertheless, these assumptions are inherent in most engineering analyses.

Next the analyst looks at those energy mechanisms which exist in his system under analysis. He then must place each mechanism into one of three categories; (1) the mechanism is judged to be unimportant and hence is excluded from the analysis, (2) the mechanism is too significient to be ignored, yet there is no need to analyze it in detail, in which case it is covered up by some artful dodge, and (3) the mechanism is significant and it is desirable to analyze it in detail. Again using the motor-generator set as an example, the magnetostrictive effect, which always exists to some degree in the laminations of the magnet, probably is insignificant and can be excluded from the analysis of the system without introducing an appreciable error. The details of what happens to the heat generated by the current flowing through the windings on the rotor may not be of significance. The total amount of energy involved in this mechanism probably is too large to be ignored so it must be included in the analysis, but not in detail, by assigning a resistance to the winding. A similar coverup job is usually done with respect to the magnetic fields established in the set by a magnet of some sort. Finally it is usually the case that there is great interest in the electrical energy supplied to and extracted from the system and in its mechanical behavior. These two energy mechanisms are then analyzed in detail; but note that this is so only because of a primary interest in the details of these mechanisms.

Having identified those mechanisms which are to be analyzed in detail, the analyst is then faced with the problem of selecting variables which are appropriate for describing these mechanisms. It is to be emphasized at this point that the information that the analyst desires is contained in these variables; hence their selection is not a matter to be treated lightly.

It is observed first that for any one mechanism two kinds of variables are required. Again citing the motor-generator set, the electrical portion of the system is usually described in terms of voltage drops and electric currents although such a selection is not mandatory. A voltage drop is typical of a class of variable which Firestone called *across variables*. Mathematicians usually call them contravariant variables. Their important characteristic is that they are defined in terms of two terminals in the system where by a terminal is meant a point or a surface needed to define the variable. Conceptually any across variable can be measured by an appropriate instrument attached to two terminals. Currents, on the other hand, are typical of a second type of variable; one which acts as if something is propagated through the system. The phrase "as if" is intended to imply, to paraphrase a comment of Faraday, "the variable may not represent the actual propagation of something through the system; but if it did, the system would act just like it does." Firestone called these *through variables* while mathematicians call them covariant variables. Any one variable of this type can be measured, at least conceptually, by breaking the system at a terminal and then inserting an appropriate meter.

Having made a selection of variables for one sort of energy mechanism, an analyst no longer has complete freedom in selecting variables for another mechanism. Note that in the illustration cited above, the product of the two variables selected to describe the electrical portion of the system, namely voltage drops and currents, has the physical dimensions of power. This fact leads to the requirement that appropriate variables for the purely mechanical portions of the system must yield the same sort of a product. Thus torques and angular velocities might be used as appropriate variables.

Having selected appropriate variables, the analyst then conceptually decomposes the system into a set of simple elements. By a simple element is meant a conceptual object, with identified terminals, whose performance is describable by a known set of relations involving the variables already selected. These relations are called *Constitutive Equations*.

Quite a lot of things can be said about elements but only a few items can be discussed here. First of all an analyst locates the terminals of the device. In this process certain points or surfaces, which seem not to be a part of the element, are needed to define some of the across variables. A very simple example arises with the rigid mass in rectilinear motion with its constitutive equation $f = m\frac{dv}{dt}$. This equation is true only if the across variable, $v$, is defined between some point on, or in, the rigid body, the center of mass for example, and an inertial reference. The inertial reference at first glance seems not to be a part of the element but this point of view is not correct. Situations analogous to the foregoing are common. Terminals which seem not to be in the element are called *External References*.

Next an analyst will seek to find a set of *independent* across variables for the element. He will try to select them so as to be the most advantageous in terms of those bits of information which he is seeking. Without going into the details, bounds can be set on the number of such variables in terms of the number of terminals on the element. If $V$ is the number of independent across variables associated with the element and $T$ is the number of terminals, then $T/2 \leq V \leq (T-1)$.

Having established a set of across variables defined as a set of terminals, a graphical symbol for the element is created by the following procedure. (1) A set of open or solid circles equal in number to the terminals on the element are put on a sheet of paper and each circle is identified with one specific terminal. (2) One line segment is drawn for each independent across variable, the line terminating on the pair of circles which correspond to the terminals of the device between which the variable is defined. (3) Certain portions of each line segment are replaced by standardized adorn-

ments which serve to show the nature of the element which stands in correspondence to the graphical symbol. Koenig has called the process of adding the adornments, *coding the symbol*. This is an apt designation. It should be borne in mind that the coding of a graphical symbol implies the constitutive equation for the element. Thus if the coding is a sawtoothed line, it is generally known that the element is a resistor with a constitutive equation $e=iR$ or $i=Ge$. The ability to write a constitutive equation in several ways is quite the rule rather than the exception.

We come now to two properties of through variables, neither of which can be proven from first principles. They represent the experience of the scientific world to date and in no case where engineering systems can be analyzed has an exception been found. First there are as many through variables associated with an element as there are across variables. Thus a single line in a graphical symbol stands in 1-to-1 correspondence with one across and one through variable. Second, the through variables act as if they are transmitted unchanged from one terminal to another along the same conceptual path suggested by the line segment. If the through variables leaving an element are considered positive and those entering negative, the foregoing continuity principle leads to the result that the algebraic sum of the through variables leaving an element is zero. This condition is aptly called the *Incidence Law*. Generally, it is stated in the simpler form; "the algebraic sum of all through variables leaving a junction point is zero." The foregoing statements show that a through variable is an oriented quantity; it can be transmitted in either of two directions.

Across variables obey a different law. Such a variable is basically, and by definition, the difference of two scalar quantities, each scalar quantity being associated with a specific terminal. The terminal associated with the scalar quantity which is subtracted is called the *Reference Terminal* for that variable. Thus every across variable is an oriented quantity in the sense that its definition must specify which of two terminals is the reference. If the algebraic sum of across variables associated with a mesh is computed in terms of the scalar quantities at terminals it is easy to show that the result is always zero. Thus across variables satisfy the Mesh Law; namely that the algebraic sum of across variables around a mesh is zero. It is well to note that whereas the Mesh Law can be proven from first principles, the Incidence Law is only a statement of experience.

At this point an analyst has in hand a set of elements each with identified terminals, an independent set of across variables, a coded graphical symbol for the element, and a set of constitutive equations. If he has done his work systematically, each element terminal corresponds to one, and only one, terminal in the original connected system. It is to be noted that whereas each element terminal corresponds to only one system terminal, which hereafter shall be called a *Junction Point*, each junction point can, and usually does, correspond to more than one element terminal. In other words, a junction point is a coincidence of several element terminals. Using these facts the analyst can generate easily a schematic diagram for the connected system by placing circles, say, on a sheet of paper; each circle corresponding to one specific junction point in the system. Graphical symbols for the elements are now connected between the junction points so that each element terminal is made coincident with the junction point to which it corresponds. This process amounts to showing graphically how the across variables satisfy the mesh laws. A more engineering like statement is that the process specifies the connectivity existing among the elements.

The foregoing process may lead to a schematic diagram with a congested, or messy, appearance. Engineers have techniques for improving the appearance of these diagrams; but of these only one is of concern to us today. This technique rests upon the fact that if it is known that the across variable defined between two supposedly different junction points is always zero, then in fact the two junction points are one and the same. Such a situation is shown in a schematic diagram by an uncoded line segment connected between the two supposedly different junction points. If desired the analyst could show that the two separate junction points are in fact the same by shrinking the uncoded line segment to zero. The process can be reversed. If a junction point has many line segments incident upon it, thus leading to a congested diagram, this single junction point can be represented in the schematic diagram as several junction points joined by uncoded line segments. Of course, this separation must be done so that all mesh relations are still satisfied and no new meshes are introduced into the diagram.

At this point the following can be asserted, "To every finite engineering system made up of elements whose constitutive equations are known, there exists a schematic diagram which specifies the connectivity of the systems and the constitutive equations for each element." The truth of this statement is obvious since a process for creating such a diagram has been defined. The principle is worded to show that if a system cannot be conceptually decomposed into elements whose terminals and constitutive equations are known then such systems can not be diagramed or analyzed at the moment. This situation exists for example with systems in which thermal energy is converted to some other form, and vice versa.

It is important to note that the mesh and the incidence laws, which are often termed the *Feld Equations*, are linear in the algebraic sense. Constitutive equations on the other hand need not be linear and the exciting situations for the future are the nonlinear ones. This difference cannot be emphasized too much for it is a rare person indeed who keeps these two items separated in his mind and in his analyses. Note that the field equations concern themselves only with the connectivity of the system; they have nothing whatsoever to do with the constitutive relations.

If one can wax philosophical for a moment, it may be pointed out that there are only two ingredients in-

volved in establishing descriptive equations for a system; these are the connectivity of the system and the constitutive equations of the elements. These two types of information correspond to the two engineering operations of selecting some devices from a bin, shelf, or ordering them from a supplier and then riveting, soldering, gluing, or otherwise joining them together. That is all there is to creating an engineering system. Thus it is seen that the generation of correct descriptive equations for a system can involve at most the connectivity of the system and the constitutive equations. Our task is to show how connectivity can be specified and how it can be made to react on the constitutive equations to yield equations for the system as a whole.

An analysis of an engineering system involves two distinct and separate phases; the formulation and the solution phase. All of the physical and engineering principles needed to deal with the system are used in the formulation phase. A bit of mathematics is used in this phase also. The more difficult mathematics is involved in the second phase, that is in solving the equations already formulated. This phase is also the most time consuming. It is a fact that generally speaking, we have mathematical tools for solving linear equations, but not for nonlinear equations. But if system equations are linear, so also are the constitutive equations for the elements. In such a situation we have linear field equations, linear constitutive equations, and linear system equations. It is not surprising therefore that these items have become confused in the minds of workers. This is not a logical position to assume and furthermore it severely restricts one's ability to deal with these exotic nonlinear engineering systems which are appearing in ever increasing numbers these latter years.

Let us now make a closer contact with the theory of linear graphs. Let a schematic diagram be given for an engineering system. In this diagram let each uncoded line segment be shrunk to zero. Finally let the coding be removed from each line segment. The result is a geometrical graph in which each vertex has a 1-to-1 correspondence to a conceptual junction point in the system and each edge, or arc, has a 1-to-1 correspondence to one across and one through variable. We now seek properties of the graph which follow the same algebraic laws as the across and through variables, i.e., the mesh and incidence laws. Again it is emphasized that these are linear laws. Let it be stated for emphasis that the linear graph has a correspondence only with the field equation of a system, not to its constitutive relations. These latter relations correspond to the coding of a schematic diagram; but all coding has been removed. It is not difficult to identify those properties of a graph that are needed. Since the theory is rather elementary and known to nearly everybody in this audience, only the highlights of the argument will be given.

Every edge in a graph has a boundary, say $\partial e_i$. Let the collection of all edge boundaries be represented by the row vector $\partial \mathbf{E}$. It is well known that the elements of $\partial E$ belong to a linear vector space. Hence any object in the space is given by a relation of the form $e_i^* = \partial E \eta$ where $\eta$ is a column vector whose elements are taken from the field of rational numbers. Furthermore an independent set of new objects can be defined by a matrix relation of the form $\partial E^* = \partial E K$ where $K$ is a nonsingular matrix with rational numbers as elements.

Each edge also provides an incidence, on two vertices. Let this be $\delta e_i$ if the edge $e_i$ is oriented away from the vertex and $-\delta e_i$ if $e_i$ is oriented toward the vertex. The collection of all edge incidences will be represented by the column vector $\delta \overline{E}$.

Each vertex can be viewed either as providing a part of the boundaries of some edges or as having an incidence because of them. The first notion for a vertex will be represented by $\hat{v}_j$ and the second by $\bar{v}_j$. Let the boundary property of vertices be collected together into the row vector $\hat{V}$ and the incidence properties into the column vector $\overline{V}$.

Since the time of Poincaré it has been known that $\partial E = \hat{V}\pi$ and $\overline{V} = \pi \delta \mathbf{E}$ where $\pi$ is a rectangular matrix. The elements of $\partial E$ are said to correspond to the elements of $\delta \mathbf{E}$. Suppose that the first relation is post multiplied by $K$ giving $\partial E K = \partial E^* = \hat{V}\pi K = \hat{V}\pi^*$. The corresponding elements from a new set of incidence objects can be found by using the principle that the incidence on any vertex is independent of the set of incidence objects used as a basis. Thus $\overline{V} = \pi \delta E = \pi(KK^{-1})\delta E = (\pi K)(K^{-1}\delta E) = \pi^* \delta E^*$ where $\delta E^* = K^{-1}\delta E$. These facts give rise to the invariance principle $\partial E \delta E = \partial E^* \delta E^*$.

A mesh is defined to be a graphical object that has no boundary. Thus a mesh is specified by a column vector $\eta$ which satisfies $\partial E \eta = \hat{V}\pi\eta = 0$. Thus $\pi\eta = 0$ for a mesh. There are an unlimited number of solutions to this equation but it is easy to show that the number of independent solutions is $\epsilon - \nu + \mathscr{S}$ where $\epsilon$ is the number of edges in the graph, $\nu$ the number of vertices and $\mathscr{S}$ the number of disjoint parts. Call this number $\mu$, the number of independent meshes.

Hereafter $K$ will be used only in the so-called canonical form in which the first $\mu$ columns define $\mu$ independent meshes. Thus $K$ can be written in the partitioned form $K = |M|P|$ where $M$, the so-called mesh connection matrix is the first $\mu$ columns. Since $K$ is nonsingular it has an inverse which can be written in the conformally partitioned form $K^{-1} = \left|\frac{Q'}{N'}\right|$ where $N$ is the so-called nodal connection matrix and the primes denote transposition. If $\partial E^*$ and $\delta E^*$ are partitioned conformally according to the scheme $\partial E^* = |\partial E_1^* | \partial E_2^*|$ and $\delta E^* = \left|\frac{\delta E_1^*}{\partial E_2^*}\right|$ it is easy to show that $\partial E = \partial E_2^* N$ and $\partial E_1^* = 0$ since the objects which give rise to the elements of $\partial E_1^*$ are all meshes.

It is clear that the boundaries of edges satisfy the mesh law. Hence using the fact that there is an across variable which corresponds to each edge it is clear that each across variable corresponds algebraically to the boundary of its corresponding edge and that $\partial E$ can be replaced by $A'$ where A is a column vector of across variables. For example $\mathrm{A} = NA_2^*$.

In a similar fashion, $\delta E$ can be replaced by a column vector T where the elements of $T$ are through variables,

but at the same time, it is necessary to impose the incidence laws. That is, to each element in $\overline{V}$ we set in correspondence a scalar, say $i_k$, and let the collection of elements be written $I$. Thus we have $I=\pi T=\pi KT^*$

$=\pi|M|P|\left|\frac{T_1^*}{T_2^*}\right|=\pi MT_1^*+\pi PT_2^*=\pi PT_2^*$ since $NM=0$

by definition. Now the elements of $I$ are set equal to zero to correspond to the incidence laws. It can be shown that $\pi P$ is nonsingular and hence that $T_2^*=0$. Since $T=KT_2^*=|M|P|\left|\frac{T_1^*}{T_2^*}\right|$ it follows that $T=MT_1^*$.

Using the results to date, it is important to observe the invariance principle yields $A'T=A_2^{*\prime}N'MT_1^*=0$ since $N'M=0$ as is easily shown from the relation $K^{-1}K=I$. This principle says that for any engineering system the formal matrix product of the across and through variables for the system is always zero. One specialized interpretation of this principle is that all of the energy supplied to a system is either stored or dissipated, a rather obvious result. It is strange that the principle is not well known since it is so basic; in fact it would appear that few people indeed have even heard of it. It is even stranger that it was only a few years ago that the principle was proved in the special case of electric networks by an engineer, at the Phillips company in Holland. Only by the argument given here has it been proved for all other cases.

Let us now consider the constitutive equations for a system. It is usually possible to write this in a matrix form like $\Lambda A=T+T_g$ where $\Lambda$ is an operator matrix and $T_g$ is a column vector of through variables delivered by the sources in the system. Now let the equation $A=NA_2^*$ be used to eliminate $A$ giving $\Lambda NA_2^*=T+T_g$. Finally let the relation be premultiplied by $N'$ giving $N'\Lambda NA_2^*=N'T+N'T_g=N'T_g$ since $N'T=N'MT_1^*=0$. This equation gives the so-called nodal formulation. It should be observed that if $\Lambda$ contains some nonlinear operators care must be taken in handling the indicated operations in order to get correct results.

A parallel type of development yields the so-called mesh formulation. Let the constitutive equations be written in the form $\psi T=A+A_g$ where $\psi$ is a matrix of operators and $A_g$ is a column vector of across variables supplied by sources. Let $T=MT_1^*$ be substituted and let the result be premultiplied by $M'$ giving $M'\psi MT_1^*$ $=M'A+M'A_g=M'A_g$ since $M'A=M'NA_2^*=0$ since $M'N=0$.

In either case the constitutive relations enter through an operator, $\Lambda$ or $\psi$, and a vector of variables specified by sources, either $T_g$ or $A_g$. The connectivity of the system is specified by $N$ in one case and $M$ in the other; the descriptions are equivalent. Thus it is seen that the connectivity of the system reacts on the constitutive equations by a congruence transformation to give a correct set of equations for the system as a whole.

If the engineering system under analysis involves only one energy mechanism which is to be analyzed in detail the foregoing procedures are adequate to handle the formulation phase. We shall let the professional mathematicians worry about the second or solution phase.

If, however, the system involves two or more energy mechanisms, each of which is to be analyzed in detail, some extra complications arise. Time will not allow a detailed discussion of this case. Since a substantial fraction of modern and interesting systems are of this type it is a bit of a pity that we cannot go into such analyses in detail. The subject is important and is worthy of a lengthy discussion on its own merits. We shall only say here that the concept of a perfect coupler, with its two types, is involved here; where by a perfect coupler is meant a conceptual mechanism which can neither create, store, nor dissipate energy but can transfer it from one part of the system to another, often changing its type in the process. These conceptual elements perform their function by imposing constraints on the variables. The imposition of these constraints on the equations obtained by the previous technique is not a hard mathematical problem, but will not be covered today. Suffice it to say that these extensions allow us to treat any system, no matter how complicated, if we know the connectivity of the system and the constitutive equations of the elements.

If most of you feel that the material I have presented today is almost insultingly simple then my lecture has been a success. I can only hope that this is the case for after all many of the important sets of concepts and procedures in the natural sciences are basically simple. Once a proper set of concepts has been assembled and extraneous matters eliminated we are left with a simple but powerful technique for making the formulation of system equations almost a routine process.

## Some Related Publications by Dr. Trent:


1. Generalized automorphisms of labelled graphs and graphic invariants, with L. Auslander, J. Math Phys. **XLII**, No. 1 (Mar. 1963).
2. A decomposition theorem concerning linear graphs, J. Math Mech. **13**, No. 2 (Mar. 1964).
3. On the realization of a linear graph given its algebraic specification, with L. Auslander, J. Acoust. Soc. Am. **33** (Sept. 1961).
4. On the topology of printed circuits, with L. Auslander, IRE Transcations CT-6 (Dec. 1959).
5. Incidence matrices and linear graphs, with L. Auslander, J. Math. Mech. **8** (Sept. 1959).
6. On the conceptual necessity and use of perfect couplers in schematic diagrams, J. Acoust. Soc. Am. **31** (Mar. 1959).
7. On the construction of schematic diagrams for mechanical systems, J. Acoust. Soc. Am. **30** (Aug. 1958).
8. Isomorphisms between oriented linear graphs and lumped physical systems, J. Acoust. Soc. Am. **27** (May 1955).
9. A note on the enumeration and listing of all possible trees in a connected linear graph, Proc. Natl. Acad. Sci. **40** (Oct. 1954).
10. An equivalent circuit for a vibrating beam which includes shear motions, J. Acoust. Soc. Am. **22** (May 1950).
11. The Laplace transform as a form of curve fitting, Am J. Phys. **17** (1949).
12. The absolute calibration of electromechanical pickups, J. Appl. Mech. (Mar. 1948).
13. Reciprocality in transducers, J. Acoust. Soc. Am. **19** (May 1947).
14. Ultrasonic measurement of wall thickness in diesel cylinder liners, J. Acoust. Soc. Am. **19** (Mar. 1947).
15. The detection of internal leaks in aircraft hydraulic systems, with R. G. Nuckolls, J. Acoust. Soc. Am. **19** (Mar 1947).

16. Some discrepancies in electromagnetic theory, Proc. Miss. Acad. Sci. (1940).
17. The concept of pressure energy, Proc. Miss. Acad. Sci. (1939).
18. The fourth dimension in electrical theory, Proc. Miss. Acad. Sci. (1939).
19. Diaphragmless microphones, Proc. Indiana Acad. Sci. (1935).
20. A study of the velocity of sound in solutions, Proc. Indiana Acad. Sci. (1935).

(Paper 69B1–136)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Character Subgroups of $F$-Groups[1]

Marvin I. Knopp[2] and Morris Newman

(November 12, 1964)

A necessary and sufficient condition is given that a subgroup of an $F$-group $G$ be definable by the vanishing of an additive character on $G$. —

1. An *F-group* is an abstract group given by a presentation of the form

$$A_1, B_1, \ldots, A_g, B_g, E_1, \ldots, E_s, P_1, \ldots, P_t;$$

$$E_1^{m_1} = \ldots = E_s^{m_s} = (\Pi\gamma_i)(\Pi E_j)(\Pi P_k) = 1,$$

where $\gamma_i = A_i B_i A_i^{-1} B_i^{-1}$. Here $m_i$ is a rational integer $\geq 2$, $s \geq 0$, $t \geq 0$, and $s+t+g > 0$. The consideration of $F$-groups arises naturally in the study of discontinuous groups, since it is known that except for certain trivial exceptions every Fuchsian group is an $F$-group when considered as an abstract group and that, conversely, every $F$-group has a faithful representation as a Fuchsian group.

If $G$ is a group, an *additive character* on $G$ is any homomorphism $\chi$ of $G$ into the additive group of the complex numbers. That is, $\chi(XY) = \chi(X) + \chi(Y)$, for $X$ and $Y$ in $G$. By a *character subgroup* of $G$ we mean any subgroup $H$ of $G$ such that $H$ is the kernel of some nontrivial additive character on $G$. Otherwise stated, there exists a nontrivial character $\chi$ on $G$ such that $H = \{X \epsilon G \mid \chi(X) = 0\}$.

Henceforth $G$ will denote an $F$-group, $\Delta$ the normal closure in $G$ of $E_1, \ldots, E_s$, and $G'$ the commutator subgroup of $G$.

The purpose of this note is to prove the following theorems.

THEOREM 1: *A subgroup* H *of* G *is a character subgroup of* G *if and only if* H *is normal in* G, $H \supset G'\Delta$, *and* G/H *has no elements of finite order.*

THEOREM 2: *A subgroup* H *of* G *is contained in a character subgroup of* G *if and only if* $H \cdot G'\Delta$ *is of infinite index in* G.

2. The essence of the proof of these theorems is contained in the two lemmas of this section.

LEMMA 1: $G/\Delta G'$ *is a free abelian group on finitely many generators.*

PROOF: A simple calculation shows that $G/\Delta G'$ is isomorphic to $(G/\Delta)/(G/\Delta)'$. (This holds for any normal subgroup $\Delta$ of $G$.) Suppose $t \geq 1$. Then eliminating the generator $P_t$ from the presentation has the concomitant effect of removing the relation $(\Pi\gamma_i)E_1 \ldots E_s P_1 \ldots P_t = 1$. Thus $G/\Delta$ is free of rank $2g+t-1$ and we can conclude that $(G/\Delta)/(G/\Delta)'$ is free abelian of rank $2g+t-1$.

If $t = 0$, then $G/\Delta$ is a fundamental group on $2g$ generators with the sole relation $\prod_{i=1}^{g} \bar{\gamma}_i = 1$, where $\bar{\gamma}_i = \gamma_i \Delta$. Hence $G/\Delta$ is isomorphic to $G^*/N$ where $G^*$ is a free group of rank $2g$ on the generators $A_i\Delta$, $B_i\Delta$, $1 \leq i \leq g$, and $N$ is the normal closure in $G^*$ of $\prod_{i=1}^{g} \bar{\gamma}_i$. Thus, $(G/\Delta)/(G/\Delta)'$ is isomorphic to $(G^*/N)/(G^*/N)'$, which in turn is isomorphic to $G^*/NG^{*\prime}$. But $N$ is contained in $G^{*\prime}$ and therefore $G^*/NG^{*\prime} = G^*/G^{*\prime}$, a free abelian group of rank $2g$.

LEMMA 2: *Let* G *be a free abelian group of finite rank* r *and let* H *be a subgroup of* G *such that* G/H *has no elements of finite order. Then* H *is a character subgroup of* G.

PROOF: Since $H$ is a subgroup of $G$, $H$ is a free abelian group of rank $s \leq r$. Suppose that

$$G = \{g_1, \ldots, g_r\}, H = \{h_1, \ldots, h_s\}.$$

Then there is an $s \times r$ rational integral matrix $A$ such that $h = Ag$, where

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_r \end{bmatrix}, \qquad h = \begin{bmatrix} h_1 \\ \vdots \\ h_s \end{bmatrix}.$$

Let $U$ be any $s \times s$ unimodular rational integral matrix and $V$ any $r \times r$ unimodular rational integral matrix. Then the elements of the vector $Uh$ are generators of $H$, the elements of the vector $Vg$ are generators of $G$, and

$$Uh = UAV^{-1} \cdot Vg.$$

Choose $U$ and $V$ so that $UAV^{-1}$ is in Smith Normal Form. Then we can write

$$Vg = \begin{bmatrix} g_1^* \\ \vdots \\ g_r^* \end{bmatrix}, \qquad Uh = \begin{bmatrix} h_1^* \\ \vdots \\ h_s^* \end{bmatrix},$$

with $h_i^* = m_i g_i^*$, for some positive integer $m_i (1 \leq i \leq s)$. Since $m_i g_i^* \epsilon H$ and $G/H$ contains no elements of finite order, it follows that $g_i^* \epsilon H$ and hence that $m_i = 1 (1 \leq i \leq s)$. That is, $h_i^* = g_i^*$, for $1 \leq i \leq s$.

[1] Research supported in part by National Science Foundation grant GP-2235.
[2] Present address: University of Wisconsin, Madison, Wis.

The desired character $\chi$ may now be defined as follows. Choose $\chi(g_i^*)=0$ for $1 \leq i \leq s$ and $\chi(g_{s+1}^*), \ldots, \chi(g_r^*)$ linearly independent over the rational numbers. It is clear that $\chi(X)=0$ for $X \epsilon H$. Furthermore, suppose that $X \epsilon G$ and that $\chi(X)=0$. Writing

$$X=c_1 g_1^* + \ldots + c_r g_r^*,$$

with rational integral $c_i$, we find that

$$\chi(X)=c_1\chi(g_1^*)+\ldots+c_r\chi(g_r^*)$$

$$=c_{s+1}\chi(g_{s+1}^*)+\ldots+c_r\chi(g_r^*)=0.$$

Since $\chi(g_{s+1}^*), \ldots, \chi(g_r^*)$ are linearly independent over the rationals, $c_{s+1}=\ldots=c_r=0$ and $X \epsilon H$. Thus $H$ is a character subgroup of $G$.

3. We are now in a position to prove Theorems 1 and 2.

*Proof of Theorem* 1. Suppose $H$ is a normal subgroup of $G$ such that $H \supset G'\Delta$ and $G/H$ has no elements of finite order. Then $(G/G'\Delta)/(H/G'\Delta) \cong G/H$ has no elements of finite order. By Lemma 1, $G/G'\Delta$ is a free abelian group, and therefore, by Lemma 2, $H/G'\Delta$ is a character subgroup of $G/G'\Delta$. The character in question on $G/G'\Delta$ can be extended to a character on $G$ in an obvious way, and it is clear that $H$ is a character subgroup of $G$.

Conversely, suppose that $H$ is the character subgroup of $G$ corresponding to the nontrivial character $\chi$. Clearly $H$ is normal in $G$ and $H \supset G'\Delta$. Suppose $X \epsilon G$ and $X^n \epsilon H$ for some positive integer $n$. Then $\chi(X^n)=n\chi(X)=0$ so that $\chi(X)=0$, and $X \epsilon H$. Thus $G/H$ has no elements of finite order, and the proof of Theorem 1 is complete.

*Proof of Theorem* 2. Suppose $H$ is a subgroup of $G$ such that $HG'\Delta$ is of infinite index in $G$. Since $[G/G'\Delta: HG'\Delta/G'\Delta]=[G: HG'\Delta]$, we conclude that $HG'\Delta/G'\Delta$ is of infinite index in the free abelian group $G/G'\Delta$. Hence $HG'\Delta/G'\Delta$ is a free abelian group on *strictly fewer* generators than the number of generators of $G/G'\Delta$. A simple calculation involving linear equations shows that there is a nontrivial character on $G/G'\Delta$ which vanishes on $HG'\Delta/G'\Delta$. Extending to a character on $G$, we obtain a nontrivial character on $G$ which vanishes on $HG'\Delta$ and *a fortiori* on $H$.

Conversely, if $H$ is a subgroup of $G$ such that a nontrivial character $\chi$ vanishes on $H$, then $HG'\Delta$ is of infinite index in $G$. For choose $X \epsilon G$ such that $\chi(X) \neq 0$. Then $\chi(X^n)=n\chi(X) \neq 0$, $(n=1, 2, 3, \ldots)$, so that $X^n \notin HG'\Delta$. This completes the proof.

(Paper 69B1–137)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# A Note on Multipliers of Difference Sets

R. A. Brualdi

(October 29, 1964)

Let $v$, $k$, $\lambda$ be integers with $0<\lambda<k<v-1$. A set $D=\{d_1, d_2, \ldots, d_k\}$ of $k$ integers distinct modulo $v$ is a *difference set* with parameters $v$, $k$, and $\lambda$ provided every nonzero residue $c$ modulo $v$ can be written in precisely $\lambda$ ways in the form $d_i-d_j\equiv c \pmod v$. An integer $t$ is a *multiplier* of $D$ provided there exists an integer $s$ such that the sets of numbers $\{td_1, \ldots, td_k\}$ and $\{d_1+s, \ldots, d_k+s\}$ coincide modulo $v$. It is shown that $-1$ cannot be a multiplier of the difference set $D$. A consequence is that a Hadamard matrix of order $v$ cannot be a symmetric circulant for $v>4$.

## Introduction

In a recent paper [1][1] Gordon, Mills, and Welch state that it is known that $-1$ cannot be a multiplier of a nontrivial difference set. No proof of this fact appears ever to have been published. The purpose of this note is to give an elementary matrix-theoretic proof of this assertion using the approach of M. Newman in [2]. Some implications are also noted.

Let $v$, $k$, $\lambda$ be integers with $0<\lambda<k<v-1$. A set $D=\{d_1, d_2, \ldots, d_k\}$ of $k$ integers distinct modulo $v$ is a *difference set* with parameters $v$, $k$, and $\lambda$ provided every nonzero residue $c$ modulo $v$ can be written in precisely $\lambda$ ways in the form

$$d_i-d_j\equiv c \pmod v,$$

with $d_i$ and $d_j$ in $D$. A simple count yields

$$k(k-1)=\lambda(v-1). \tag{1}$$

An integer $t$ is a *multiplier* of $D$ provided there exists an integer $s$ such that the sets of numbers $\{td_1, td_2, \ldots, td_k\}$ and $\{d_1+s, d_2+s, \ldots, d_k+s\}$ coincide modulo $v$.

Let $P$ be the $v$-square permutation matrix

$$P=\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ . & . & . & \ldots & . \\ 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

and $C$ the circulant matrix

$$C=\sum_{i=1}^{k} P^{d_i} \tag{2}$$

We have $P^v=I$, and the properties of the difference set $D$ imply that

$$CC^T=C^TC=(k-\lambda)I+\lambda J, \qquad CJ=JC=kJ \tag{3}$$

where $J$ is the $v$-square matrix of all 1's. For $t$ an integer let

$$C_t=\sum_{i=1}^{k} P^{td_i}.$$

Then $t$ is a multiplier of the difference set $D$ if and only if there exists an integer $s$ where $0\leqq s\leqq v-1$ such that

$$C_t=P^sC. \tag{4}$$

THEOREM: *Let* $D=\{d_1, d_2, \ldots, d_k\}$ *be a difference set with parameters* $v, k,$ *and* $\lambda$ *where* $0<\lambda<k<v-1$. *Then* $-1$ *is not a multiplier of* $D$.

PROOF: Suppose $-1$ is a multiplier. Then

$$C_{-1}=\sum_{i=1}^{k} P^{-d_i}=C^T,$$

and thus

$$C^T=P^sC \tag{5}$$

for some integer $s$, $0\leqq s\leqq v-1$. If $s$ is even, say $s=2r$, then (5) can be written as $(P^rC)^T=P^rC$ where $P^rC$ also satisfies (3). If $s$ is odd and $v$ is odd, then since $P^{v+s}=P^v\cdot P^s=P^s$ a similar remark holds. In case $s$ is odd, say $s=2r+1$, and $v$ is even then (5) can be written as $(P^rC)^T=P(P^rC)$ where $P^rC$ again satisfies (3). Hence we may assume that either

$$C^T=C, \; v \text{ odd or even}, \tag{6}$$

or

$$C^T=PC, \; v \text{ even}. \tag{7}$$

We proceed to show that a circulant matrix $C$ of 0's and 1's which satisfies (3) with $0<\lambda<k<v-1$ and either (6) or (7) does not exist. Let

$$c_0, c_1, c_2, \ldots, c_{v-1}$$

denote the first row of $C$, so that

$$c_0, c_{v-1}, c_{v-2}, \ldots, c_1$$

is the first row of $C^T$.

[1] Figures in brackets indicate the literature references at the end of this paper.

Case 1. $C^T=C$; $v$ even, say $v=2r$.

Comparing the first rows of $C^T$ and $C$ we find that $c_j=c_{r-j}$, $j=1, 2, \ldots, v-1$. Thus the matrix $C$ assumes the form

$$C=\begin{bmatrix} c_0 & c_1 & c_2 & \ldots & c_{r-1} & c_r & c_{r-1} & \ldots & c_2 & c_1 \\ c_1 & c_0 & c_1 & \ldots & c_{r-2} & c_{r-1} & c_r & \ldots & c_3 & c_2 \\ c_2 & c_1 & c_0 & \ldots & c_{r-3} & c_{r-2} & c_{r-1} & \ldots & c_4 & c_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

By (3) the inner product of any two distinct rows of $C$ is equal to $\lambda$. The inner product of row 1 and row 2 shows that $\lambda$ is even. The inner product of row 1 and row 3 is congruent to $c_1^2+c_{r-1}^2=c_1+c_{r-1}$ modulo 2. Since $\lambda$ is even, $c_1=c_{r-1}$. The inner product of row 1 and row 5 gives $c_2=c_{r-2}$. Continuing in this way we conclude that the first row of $C$ has the form

$$c_0, c_1, c_2, \ldots, c_2, c_1, c_r, c_1, c_2, \ldots, c_2, c_1.$$

The inner product of row 1 and row $(r+1)$ then yields

$$\lambda=c_1^2+c_2^2+\ldots+c_2^2+c_1^2+2c_0c_r+c_1^2+c_2^2+\ldots+c_2^2+c_1^2$$

or since the $c$'s are either 0 or 1,

$$\lambda=c_1+c_2+\ldots+c_2+c_1+2c_0c_r+c_1+c_2+\ldots+c_2+c_1.$$

Since the sum of the entries in the first row of $C$ is equal to $k$, we then have $\lambda-k=2c_0c_r-c_0-c_r$. Because $c_0$ and $c_r$ are either 0 or 1, this implies $\lambda=k$ or $\lambda=k-1$. Since $\lambda=k-1$ implies by (1) that $k=v-1$, we have a contradiction.

Case 2. $C^T=C$; $v$ odd, say $v=2r+1$.

In this case the matrix $C$ has the form

$$\begin{bmatrix} c_0 & c_1 & c_2 & \ldots & c_{r-1} & c_r & c_r & c_{r-1} & \ldots & c_2 & c_1 \\ c_1 & c_0 & c_1 & \ldots & c_{r-2} & c_{r-1} & c_r & c_r & \ldots & c_3 & c_2 \\ c_2 & c_1 & c_0 & \ldots & c_{r-3} & c_{r-2} & c_{r-1} & c_r & \ldots & c_4 & c_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

The inner product of row 1 and row 2 shows that $\lambda \equiv c_r \pmod 2$.

The inner product of row 1 and row 3 shows that $\lambda \equiv c_1 \pmod 2$. Continuing in this way we see that

$$\lambda \equiv c_j \pmod 2, \qquad j=1, 2, \ldots, r.$$

Since the $c$'s are either 0 or 1, we conclude that $c_1=c_2=\ldots=c_r$. Since by (3) $k=c_0+2(c_1+\ldots+c_r)$ and $k \geqq 2$, this common value must be 1. But then $k=v$ or $k=v-1$ and we have a contradiction.

Case 3. $C^T=PC$, $v$ even, $v=2r$.

Comparing the first rows of $C^T$ and $PC$, we see that $c_j=c_{r-1-j}$, $j=0, 1, \ldots, v-1$. The matrix $C$ then has the form

$$\begin{bmatrix} c_0 & c_1 & c_2 & \ldots & c_{r-1} & c_{r-1} & \ldots & c_2 & c_1 & c_0 \\ c_0 & c_0 & c_1 & \ldots & c_{r-2} & c_{r-1} & \ldots & c_3 & c_2 & c_1 \\ c_1 & c_0 & c_0 & \ldots & c_{r-3} & c_{r-2} & \ldots & c_4 & c_3 & c_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

The inner product of row 1 and row 3 shows $\lambda$ must be even. The inner product of row 1 and row 2 implies $\lambda \equiv c_0+c_{r-1} \pmod 2$. Hence $c_0=c_{r-1}$. The inner product of row 1 and row 4 yields $c_1=c_{r-2}$. In general, $c_j=c_{r-1-j}$, $j=0, 1, \ldots, r-1$. Hence the first row of $C$ has the form

$$c_0, c_1, \ldots, c_1, c_0, c_0, c_1, \ldots, c_1, c_0.$$

The inner product of row 1 and row $(r+1)$ then gives

$$\lambda=c_0+c_1+\ldots+c_1+c_0+c_0+c_1+\ldots+c_1+c_0=k.$$

But $\lambda=k$ is a contradiction. This completes the proof of the theorem.

If $t$ is a multiplier of a difference set $D$ with parameters $v$, $k$, and $\lambda$, then $(t, v)=1$. Moreover it is easily verified that the set of multipliers modulo $v$ form a multiplicative group $M$ which is a subgroup of the multiplicative group $G$ of all integers modulo $v$ which are prime to $v$. Let $|M|$ denote the order of $M$ and $\varphi(v)$ the Euler function of $v$, that is the number of positive integers not greater than and prime to $v$. Then

COROLLARY 1: M *is a proper subgroup of* G *and hence* $|M| \leqq \frac{\varphi(v)}{2}$.

PROOF: $-1$ is an element of $G$ but by the theorem not of $M$.

A $(v, k, \lambda)-$*matrix* is a $v$-square matrix of 0's and 1's which satisfies

$$AA^T=(k-\lambda)I+\lambda J,$$

where $0<\lambda<k<v-1$. In the proof of the theorem we have shown that there do not exist $(v, k, \lambda)-$matrices $A$ which are circulants and satisfy $A^T=P^sA$ where $0 \leqq s \leqq v-1$. In particular there do not exist symmetric circulant $(v, k, \lambda)-$matrices.

We are indebted to Morris Newman for pointing out to us the following consequence. A *Hadamard matrix* $H$ is a $v$-square matrix of $+1$'s and $-1$'s such that $HH^T=vI$. It has been conjectured [3] that a Hadamard matrix cannot be a circulant for $v>4$. In this regard we have

COROLLARY 2: *A Hadamard matrix cannot be a symmetric circulant for* $v>4$.

PROOF: For, if $H$ is a symmetric Hadamard circulant, then

$$K=\frac{1}{2}(H+J)$$

is a symmetric circulant $(v, k, \lambda)$−matrix with $k = \frac{v \pm \sqrt{v}}{2}$ and $\lambda = \frac{v \pm 2\sqrt{v}}{4}$, unless $k = 0, 1, v-1, v$.

But it is easily checked that these requirements cannot be satisfied for $v > 4$. This proves the corollary.

---

Part of the work for this paper was done while the author was a National Research Council–National Bureau of Standards Postdoctoral Research Associate, 1964-1965. The author wishes to thank Morris Newman for helpful comments concerning the exposition of this paper.

Note added in proof: In the Canadian Journal of Mathematics (Vol. **16**, 1964) E. C. Johnsen in his paper "The Inverse Multiplier For Abelian Group Difference Sets" investigates a problem similar to ours for the more general abelian group difference sets. In the then special case of our circumstances he obtains by nonelementary means a proof of the theorem in our paper.

## References

[1] B. Gordon, W. H. Mills, and L. R. Welch, Some new difference sets, Can. J. Math. **14**, 614-625 (1962).

[2] M. Newman, Multipliers of difference sets, Can. J. Math. **15**, 121-124 (1963).

[3] H. J. Ryser, Combinatorial Mathematics, Carus Math. Monograph, No. 14, (John Wiley & Sons, Inc., New York, N.Y., 1963).

(Paper 69B1–138)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Modification of Edmonds' Maximum Matching Algorithm*

C. Witzgall and C. T. Zahn, Jr.**

(November 16, 1964)

Edmonds developed an efficient algorithm for finding in a given graph $G$ a matching of maximum cardinality. This algorithm "shrinks" parts of the graph $G$. Although helpful to the intuitive understanding of the theory, shrinking is complicated to implement on an electronic computer. The modification presented in this paper avoids shrinking. It employs instead a treelike arrangement of alternating paths. The possibility of such an arrangement is also of theoretical interest, and its proof forms the main part of the paper.

## 1. Introduction

A matching $M$ in a graph $G$ is a set of edges of $G$ such that no two meet the same vertex. Edmonds [1][1] developed an efficient algorithm for finding in a given graph $G$ a matching of maximum cardinality. The maximum matching problem belongs to a class of problems which for the most part have defied efficient treatment. These are integer linear programming problems—in particular those associated with networks. A celebrated and still largely intractable representative of this class is the traveling salesman problem.

The techniques on which Edmonds bases his algorithm have some interesting theoretical consequences. For example, Edmonds assigns the labels "outer" and "inner" to the vertices of the graph $G$. This assignment is based on the underlying matching; however, it turns out to be the same for all maximum matchings. Therefore the "inner" and "outer" property is an invariant of the graph itself. The outer vertices (with respect to any maximum matching) are precisely those vertices which are "exposed"—that is, they meet no matched edge—in some maximum matching of the graph. The set of outer vertices is of considerable structural interest (compare Edmonds [1]).

The notion of outer vertices is closely related to accessibility by simple alternating paths. To clarify this statement, we need some definitions.

First we note that it involves no loss of generality when searching for a maximum matching to restrict our considerations to graphs whose edges can be conceived of as unordered pairs of vertices. Indeed, if the graph $G$ possesses pairs of vertices that are connected by a bundle of more than one edge, we construct a subgraph $\hat{G}$ by deleting from every such bundle all but one of its edges. Every maximum matching of $\hat{G}$ is also a maximum matching of $G$.

Due to the absence of multiple edges, we may characterize *paths* by sequences of vertices $(v_1, \ldots, v_n)$ every two consecutive ones of which are adjacent. A path is *simple* if its vertices, and therefore its edges, are distinct. A *closed path* is *simple* if its vertices and its edges are distinct with the only exception $v_1 = v_n$. Every path *forms* a subgraph consisting of the vertices $v_i$ of the path and the edges joining its successive vertices. A simple closed path, for instance, forms a *circuit*, that is, a connected graph each of whose vertices is incident to exactly two vertices.

A graph $G$, along with a matching $M$, will be called a *matched graph* $(G, M)$. The set of edges of $G$ not in $M$ will be denoted by $\overline{M}$. An *alternating* path in $(G, M)$ is a path whose successive edges are alternately in $M$ and $\overline{M}$. In particular, the path of one node is an alternating path. An *exposed* vertex in $(G, M)$ is one which is incident to no edges in $M$. An *augmenting path* is a simple alternating path connecting two exposed vertices. The significance of augmenting paths for matchings is demonstrated by the theorem of Berge [4]: *A matching* M *is of maximum cardinality if and only if there exists no augmenting path in* (G, M). The easier part of this theorem is that, in the presence of an augmenting path, a matching cannot be maximum. Indeed, by changing all nonmatched edges of an augmenting path into matched edges and vice versa, the cardinality of the matching is increased.

According to the theorem of Berge, the maximum matching problem can be solved by searching for all alternating paths from each exposed vertex. However, one can do better. To see this we reformulate Berge's theorem, defining a vertex $v$ to be an *outer* vertex *rooted* at $e$ if $e$ is an exposed vertex which is linked to $v$ by a simple alternating path of even length (= number of edges). (In particular, we regard all exposed vertices as outer.) The reason for considering outer vertices becomes evident if one examines an augmenting path connecting two exposed vertices $e_1$ and $e_2$. Let $v_1$ and $v_2$ be the neighbors of $e_1$ and $e_2$ within the augmenting path. Then $v_2$ is an outer

* Supported by the Army Research Office (Durham).

** Present address: General Electric Computer Department, 679 Leesburg Pike, Falls Church, Va.

[1] Figures in brackets indicate the literature references at the end of this paper.

vertex rooted at $e_1$, and $v_1$ is an outer vertex rooted at $e_2$. Therefore, *a matching is maximum if and only if no exposed vertex* e *is adjacent to an outer vertex which is rooted at an exposed vertex different from* e.

For the purpose of establishing maximality or nonmaximality it is therefore sufficient to search for all outer vertices. This is a genuine improvement over searching for all alternating paths, since there are in general more alternating paths emanating from exposed points than there are outer vertices. Moreover, a list of vertices is easier to handle and store than a list of paths.

The method of Edmonds, as well as its modification described in this paper, searches for outer vertices. It should be realized, however, that if the matching turns out to be not maximum, that is, if there are outer vertices adjacent to exposed vertices other than their roots, then we still face the problem of actually finding an augmenting path. This means essentially that we should be able to retrieve for each outer vertex a simple alternating path that leads back to a root. The difference between Edmonds' method and our modification lies in the kind of additional information furnished to permit this "back-tracing".[2] Edmonds shrinks the graph in such a manner that at each outer vertex in the shrunken graph there is only one possibility for stepping backwards along an alternating path: one is forced to proceed in the right direction. The modification presented in this paper is based on the notion of a "predecessor." It should be noted that in general the two methods will find different back-paths. When expanded into paths of the original graph, the back-paths defined by Edmonds' method will in general not display the treelike arrangement described in the next section.

[2] Rather surprisingly, in both methods the process of finding the outer vertices also requires back-tracing from outer vertices already found. The importance of back-tracing is therefore not restricted to augmentation. This "back-tracing" should not be confused, however, with so-called "back-tracking" methods that usually involve an unpleasant amount of trial and error. Our algorithm, as well as Edmonds', is a good one.

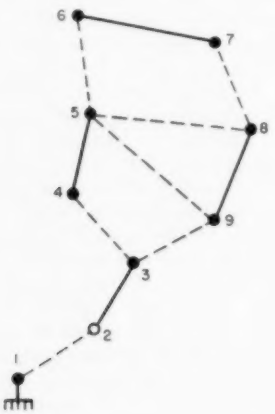



FIGURE 1. $p(3)=1$, $p(5)=3$, $p(7)=5$, $p(9)=7$, $p(8)=3$, $p(6)=3$, $p(6)=8$, $p(4)=6$.
——, Matched edge.
●, Outer points.

We assign to each nonexposed outer vertex $v$ another outer vertex $u=p(v)$ which is connected to $v$ by an alternating path of length two. Vertex $u=p(v)$ is called the predecessor of $v$. We shall prove that this assignment of predecessors can be specified in such a way that the alternating path formed by consecutive predecessors is simple and joins $v$ to an exposed vertex. (This is what we mean in saying a complete set of alternating back-paths has been arranged in a tree-like manner.) We employ the fact that an alternating path is uniquely determined if only every other vertex is given.

As an example consider the graph in figure 1. The vertices are numbered *1*, . . . , *9*. Heavy lines distinguish the matched edges. Vertex *1* is the only exposed vertex. As an augmenting path would require two exposed vertices, the matching is maximum. Solid points designate outer vertices. Vertex *2* is not outer, since every alternating path of even length joining vertex *2* to the "root", as for instance (*2*, *3*, *9*, *8*, *5*, *4*, *3*, *2*, *1*), passes the edge (*2*, *3*) twice, thus being not simple.

The assignment of predecessors $p(v)$, as specified in figure 1, yields a simple back-path for each outer vertex. Vertex 7, for instance, is joined to vertex *1* by the path (*7*, *6*, $5=p(7)$, 4, $3=p(5)$, 2, $1=p(3)$).

## 2. Graph of Bi-edges

The situation is best understood by considering the *graph of bi-edges* (Edmonds [5]) which is associated with the matched graph $(G, M)$. A *bi-edge* is an alternating path of length two, that is, a path consisting of an unmatched edge followed by a matched one. Bi-edges are the units of which even-length alternating paths are composed. The bi-edge graph $B(G, M)$ is defined over the same vertex set as $G$, two vertices being linked if they are the end-points of a bi-edge in $(G, M)$. We regard bi-edges as directed: they lead from the end that is incident with the matched part to the end that is incident with the unmatched part of the bi-edge (fig. 2). Consequently, the graph of bi-edges is a "directed" graph. A *directed graph* consists of vertices, and edges which are ordered pairs of the vertices. According to this definition, a directed graph can have two edges, $(v_1, v_2)$ and $(v_2, v_1)$, incident to the same pair of vertices. Figure 3 shows the graph of bi-edges that corresponds to the matched graph in figure 1.

A path in a directed graph can be characterized again by a sequence of vertices $(v_1, v_2, \ldots, v_n)$. However, we insist that each linking edge lead from a vertex to its successor in the path; more precisely, we require the edge that links two successive vertices $v_i$, $v_{i+1}$ to be the edge $(v_i, v_{i+1})$ in this order. Each path in the bi-edge graph $B(G, M)$ corresponds to a unique alternating path in $(G, M)$. We shall say that a path in $B(G, M)$ *expands* into the corresponding alternating path in $(G, M)$. Alternating paths gained by expansion of bi-edge paths have even length (= number of edges, where each edge counts as often as it is traversed).

FIGURE 2

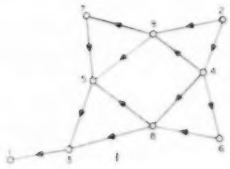


FIGURE 3

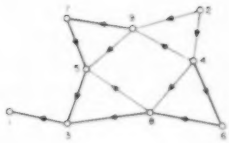


FIGURE 4

On the other hand, each alternating path of even length corresponds to a path in the bi-edge graph. However, there is one crucial difficulty: *a simple path in B(G, M) may expand into an alternating path with double edges.* For an example consider the matched graph in figure 1. Its graph of bi-edges is given in figure 3. The simple path (*2, 9, 5, 3, 1*) in the bi-edge graph expands into the alternating path (*2, 3, 9, 8, 5, 4, 3, 2, 1*) with the edge (*2, 3*) occuring twice.

In view of this situation we call a path in the bi-edge graph *legal* if it expands into a simple alternating path. Vertices that can be reached from a vertex *r* by legal paths are called *legally accessible* from *r*. We then have the

THEOREM. *Let* B(G, M) *be the graph of bi-edges of a matched graph* (G, M), *let* r *be an arbitrary vertex of* B(G, M), *and denote by* Ω *the set of all vertices of* B(G, M) *that are legally accessible from* r. *Then* B(G, M) *contains as a subgraph a tree* T *which is rooted at* r, *such that* (*i*) T *has* Ω *as its vertex set, and* (*ii*) *every simple path in* T *that joins a vertex* v *to the root* r *is legal.*

Note that the theorem would be trivial if the legality requirement were dropped.

Any rooted tree $T$ in $B(G, M)$ corresponds to an assignment of predecessors as discussed before. For instance, the predecessors specified in figure 1. define the subtree shown by heavy lines in figure 4.

The relation between the above theorem and the algorithm for maximum matchings that will be proposed in this paper is as follows: generality of the theorem is not restricted by assuming that the root *r* is an exposed vertex. Indeed, if *r* meets a matched edge, then removing this edge from the matching does not change the set of simple alternating paths emanating from *r* along nonmatched edges. The problem then becomes to arrange into a tree $T \subseteq B(G, M)$ alternating paths which connect outer vertices to an exposed vertex *r*. Our *algorithm* $\mathfrak{T}$ will sucessively construct this tree, thereby proving the theorem. On the other hand, the algorithm will rely at each step on the fact that the alternating paths which connect the outer vertices established at this moment to the root *r* are properly arranged in a tree-like manner. A combination of the algorithm $\mathfrak{T}$ with augmentation using augmenting paths gives the *algorithm* $\mathfrak{M}$, which yields a maximum matching.

## 3. Labeled Subgraphs

Let *r* be an exposed vertex in a matched graph $(G, M)$. We define a "*labeled subgraph*" with the root *r* to be a quadruple

$$L = (A, \Omega, I, p)$$

where $A$ is a subgraph of $G$. The vertices of $A$ are labelled either "outer" or "inner", $\Omega$ denoting the set of vertices labeled "outer", and $I$ the set of vertices labeled "inner". $\Omega$ and $I$ are disjoint. Finally, $p$ is a single-valued function $p\colon \Omega - \{r\} \to \Omega$ called the *predecessor function.*

In addition, we shall require certain relations to hold between the elements of the quadruple $(A, \Omega, I, p)$. To formulate these requirements, and for later use, we introduce the following *matching function m*, which maps the set of all nonexposed vertices of $G$ into itself, and which is defined by

$$m(v) := u \qquad \text{if and only if} \qquad (v, u) \in M.$$

The requirements for a quadruple $(A, \Omega, I, p)$ to be a labeled subgraph then are:

(i) *If one vertex of a matched edge is in* A *then the entire edge is in* A.

(ii) *If* x ∈ I *then* m(x) ∈ Ω

(iii) *The edge* (m(v), p(v)) *is in* A *for every* v ∈ Ω − {r}

Note that (ii) and (iii) imply that $I$ and $\Omega - \{r\}$ consist of nonexposed vertices. It also follows that for $x \in I$, $(x, pm(x))$ must be in $A$.[3]

(iv) *Decomposition: If* x ∈ I *then deleting the edge* (x, pm(x)) *decomposes* A *into two connected components.*

(v) *Legality: For each vertex* v ∈ Ω *the sequence of vertices* (v, p(v), $p^2$(v), . . .) *eventually contains* r, *and the back-path*

$$P(v) := (v, m(v), p(v), mp(v), p^2(v) \ldots, p^n(v) = r)$$

*has no double vertices.*

The back-path $P(v)$ lies entirely in $A$ according to (i) and (ii).

The vertices of the form $p^k(v)$, including the vertex $v$ itself, are said to be *even* in $P(v)$, the rest being *odd*. The even vertices, in particular the root $r$, are of course in Ω. The sequence of even vertices in a back-path defines a simple and legal path in the graph of bi-edges $B(G, M)$.

By virtue of the recursive structure of the back-paths, we have $P(u) \subseteq P(v)$ if $u$ is even in $P(v)$. Therefore we may define a partial ordering in Ω by letting:

$$u \leqslant v \text{ if and only if } u \text{ is even in } P(v).$$

Any two vertices $u$ and $v$ in Ω have at least one common lower bound. By finiteness, they have greatest common lower bounds. Let $w$ and $\overline{w}$ be such greatest common lower bounds. Since $w \leqslant u$ and $\overline{w} \leqslant u$, both $w$ and $\overline{w}$ are even in $P(u)$, and therefore comparable. Hence $\overline{w} = w$ because they are both maximal. It follows that the greatest common lower bound $w = u \cap v$ is unique.

We shall now describe two operations which enlarge labeled subgraphs. If $L$ is a labeled subgraph, and there is a bi-edge $(v, x, u)$ with $u \in \Omega$, $x \notin \Omega \cup I$, $(v, x) \in M$, then $v \in \Omega \cup I$ by (i). We obtain a new quadruple $\tilde{L} = (\tilde{A}, \tilde{\Omega}, \tilde{I}, \tilde{p})$ by letting $\tilde{A} := A \cup (v, x) \cup (x, u)$, $\tilde{\Omega} := \Omega \cup \{v\}$, $\tilde{I} := I \cup \{x\}$, and

$$\tilde{p}(z) := \begin{cases} p(z) \text{ for } z \in \Omega \\ u \text{ for } z = v. \end{cases}$$

In other words, we add the bi-edge $(v, x, u)$ to $A$, labeling $v$ "outer" and $x$ "inner", and putting $\tilde{p}(v) := u$. The predecessor function is not changed for the old Ω. Clearly, $L$ is a labeled subgraph. We write

$$\tilde{L} := F(L)$$

to indicate that $\tilde{L}$ is obtained by a *forward step.*

If $L$ is a labeled subgraph, and there is an edge $(s, t)$ which is not in $A$ but joins two vertices of Ω, then again we can enlarge the subgraph: Since $s$ and $t$ belong to Ω, and $L$ is legal, we have back-paths $P(s)$ and $P(t)$. Let $b := s \cap t$ denote the greatest common lower bound of $s$ and $t$ with respect to the ordering defined above. Then $P(s) \supseteq P(b)$ and $P(t) \supseteq P(b)$. The paths $P(s) - P(b)$ and $P(t) - P(b)$ together with $b$ form a path $P(s, b, t)$ joining $s$ and $t$. Following Edmonds [1], we call this path a *blossom* with the *base* $b$ (see fig. 5). We shall see later that all vertices of a blossom are indeed legally accessible: the inner vertices, for instance in $P(s) - P(b)$, are joined to $t$ by an even alternating path, which will be shown to be simple. Labeling the inner vertices in a blossom "outer", and adjoining $(s, t)$ to $A$ is called a *blossom step*[4]

$$\tilde{L} := B(L).$$

The details of a blossom step are as follows: Let $\overline{s}$ be defined as the greatest even vertex in $P(s)$ such that $P(\overline{s}) - P(b)$ contains no inner vertex, and let $\overline{t}$ be similarly defined. Either $\overline{s} = s$, or the vertex $x \in P(s) - P(\overline{s})$ which is next to $\overline{s}$ in $P(s)$ is in $I$. If $\overline{s} \neq s$, then let $u_1 := x$, and let $u_k$ be defined inductively as the unique vertex of $P(s)$ such that $pm(u_k) = m(u_{k-1})$ until some $u_m = m(s)$. Put the set $\{u_i \mid i = 1, \ldots, m\}$ into Ω, deleting vertices from $I$ whenever necessary, and define the new predecessor function by $\tilde{p}(u_i) = u_{i+1}$ for $1 \leqslant i \leqslant m$ where $u_{m+1} = t$. Notice that some $u_i$ may be in Ω and already have a predecessor which is hereby changed. The same procedure is repeated for $t$ in place of $s$, and then we add the edge $(s, t)$ to $A$ to obtain the new graph $\tilde{A}$. If the counterparts of the $u_i$ in the path $P(t)$ are denoted $v_j$ then formally we get $\tilde{A} := A \cup (s, t)$, $\tilde{\Omega} := \Omega \cup \{u_i\} \cup \{v_j\}$, $\tilde{I} := I - \{u_i\} - \{v_j\}$,

$$\tilde{p}(z) := \begin{cases} u_{i+1} \text{ for } z = u_i \\ v_{j+1} \text{ for } z = v_j \\ p(z) \text{ otherwise.} \end{cases}$$

The above definition of $\tilde{p}(z)$ implicitly requires the sets $\{u_i\}$ and $\{v_j\}$ to be disjoint. Since the $u_i$ are

[3] We write $pm(x)$ for $p(m(x))$, $mp(x)$ for $m(p(x))$, $p^2(x)$ for $p(p(x))$.

FIGURE 5

[4] In practice one will use a more elaborate blossom step. Let $x$ and $y$ be two adjacent vertices of the blossom $P(s, b, t)$. After execution of the blossom step defined by $s$ and $t$, both $x$ and $y$ will be outer whereas their connecting edge $(x, y)$ need not be in $A$. In this case, $x$ and $y$ will give rise to another blossom-step. The corresponding blossom, however, is easily verified to be contained in the circuit formed by $P(s, b, t)$ and the edge $(s, t)$. This blossom therefore contains no inner vertices, and all the blossom step does is adjoin the edge $(x, y)$ to $A$. The more elaborate blossom step consists of adding not only the edge $(s, t)$ but all other "diagonals" of the blossom $P(s, b, t)$ to $A$.

odd in $P(s)$ and the $v_j$ are odd in $P(t)$, this will be a consequence of the following

LEMMA 1. *Let* s *and* t *be vertices in* $\Omega$, *and put* $b := s \cap t$. *If the vertex* v *is of equal parity in both* P(s) *and* P(t), *then* $v \in P(b)$.

PROOF: If $v$ is even in both $P(s)$ and $P(t)$, then $v \leq s$ and $v \leq t$. Hence $v \leq s \cap t = b$, which implies $v \in P(b)$ by definition of the partial ordering.

If $v$ is odd in both $P(s)$ and $P(t)$, then $v$ cannot be the root $r$, because $r$ is even in both $P(s)$ and $P(t)$. Hence $m(v)$ is defined, and it is even in both $P(s)$ and $P(t)$. By the preceding argument we then conclude $m(v) \in P(b)$ and therefore $v \in P(b)$.

Before proving that the blossom step $B$ as defined above always yields a new labeled subgraph $\bar{L}$, we illustrate the blossom step by two examples.

The first example (fig. 5) shows that the two halfs of the blossom, $P(s) - P(b)$ and $P(t) - P(b)$, need not be disjoint.

Starting with the root $r = 1$, the assignment of predecessors has progressed as follows:

$p(3) = 1$ forward step
$p(5) = 3$ forward step
$p(7) = 5$ forward step
$p(9) = 7$ forward step
$p(8) = 3$, $p(6) = 8$, $p(4) = 6$ } blossom step (9, 3), base 3
$p(10) = 9$ forward step

Now a blossom step is due for $s = 10$ and $t = 4$. The back-paths are

$$P(s) = (10, 11, 9, 8, 7, 6, 5, 4, 3, 2, 1)$$
$$P(t) = (4, 5, 6, 7, 8, 9, 3, 2, 1),$$

$b = 3$ is the base. It is seen that the part (4, 5, 6, 7, 8, 9) of the blossom is traversed by $P(s)$ as well as $P(t)$, however in opposite directions. Changing predecessors in the area common to $P(s)$ and $P(t)$ would clearly destroy legality. This is one reason for changing predecessors only up to the last inner vertex. Later on we shall indeed see that $P(s) - P(\bar{s})$ and $P(t) - P(\bar{t})$ are always disjoint.

Even if the blossom has no double points, legality may be destroyed if predecessors are changed beyond the last inner vertex. This is shown by our second example (fig. 6).

Starting with the root $r = 1$, the assignment of predecessors has progressed as follows

$p(3) = 1$ forward step
$p(5) = 3$ forward step
$p(4) = 1$, $p(2) = 4$ } blossom step (5, 1), base 1
$p(7) = 2$ forward step
$p(8) = 7$ forward step
$p(9) = 7$ blossom step (8, 7), base 7.

Again a blossom step is due for $s = 9$ and $t = 4$. The

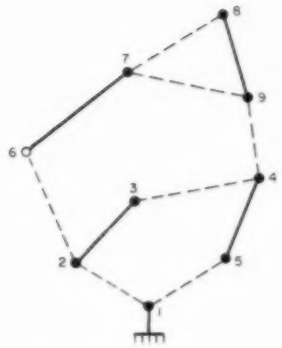



FIGURE 6. ——, *Matched edges;* ●, *outer points.*

back-paths are

$$P(s) = (9, 8, 7, 6, 2, 3, 4, 5, 1)$$
$$P(t) = (4, 5, 1),$$

yielding $b = 4$ as base. The blossom step as defined before puts

$$\tilde{p}(8) = 4$$
$$\tilde{p}(6) = 8.$$

Note that the predecessor of 8 has been changed from 7 to 4. If one would change the predecessor beyond the inner point 6, that is, if one puts

$$\tilde{p}(3) = 6$$

then the new back-path $\tilde{P}(5)$ will not be simple.

## 4. Preservation of Legality

It must be shown that $F(L)$ and $B(L)$ are again labeled subgraphs. Verifying (i) to (iii) can be left to the reader. Proving the decomposition property (iv) and, in particular, the legality (v) is the main burden of this paper.

LEMMA 2. *The decomposition property (iv) is preserved by the operations* F *and* B.

PROOF: In the case of a forward step $F$ there is really nothing to prove since no new circuits are introduced in $A$. In the case of a blossom step, however, new circuits are generated, and we have to show that these circuits contain no inner vertices.

Let then $s$ and $t$ be two adjacent outer vertices determining a blossom $P(s, b, t)$ with $b := s \cap t$, and execute the corresponding blossom step. Assume that the new subgraph $\bar{A}$ contains a circuit $C$ with a vertex $x \in \bar{I}$. Since $L$ satisfies the decomposition property (iv) and since $x \notin \Omega$, $C$ cannot be a circuit in $A$, and must therefore contain the new edge $(s, t)$. Replacing this edge by the blossom path $P(s, b, t)$ yields a closed path $C^*$ (not necessarily a circuit), which lies entirely in the old subgraph $A$, and which still contains the inner

vertex $x$. Since $P(s, b, t) \cap \tilde{I} = \phi$ whereas $x \in \tilde{I}$, the vertex $x$ is not a double point of $C^*$. Hence $X$ is contained in a subcircuit [5] of $C^* \subseteq A$, contradicting (iv).

LEMMA 3. *The legality property (v) of a labeled subgraph* L *is preserved by the operations* F *and* B.

PROOF: In the case of forward steps $F$, the proof presents no difficulties, and may be left to the reader. Let then $s$ and $t$ be again two adjacent vertices in $\Omega$ such that the edge $(s, t)$ is not in $A$, and execute the corresponding blossom step $L := B(L)$. Define for $v \in \tilde{\Omega}$ the – possibly infinite – sequence

$$\tilde{P}(v) := (v, m(v), \tilde{p}(v), m\tilde{p}(v), \tilde{p}^2(v), \ldots).$$

It is to be shown that $\tilde{P}(v)$ is a simple alternating path connecting $v$ to the root $r$. Trivially, this is the case if $\tilde{P}(v) = P(v)$. Otherwise, let $w$ be the even vertex in $\tilde{P}(v)$ that is closest to $v$, and that is odd in $P(s) - P(\bar{s})$ or $P(t) - P(\bar{t})$. Since predecessors are not altered outside the set of odd vertices of $\{P(s) - P(\bar{s})\}$ and the set of odd vertices of $\{P(t) - P(\bar{t})\}$, we have

(vi) $$\tilde{P}(v) - \tilde{P}(w) = P(v) - P(w).$$

Without loss of generality we may assume that $w \in P(s) - P(\bar{s})$. It follows again by the definition of the blossom step $B$ that the sequence $(v, \tilde{p}(v), \tilde{p}^2(v), \ldots)$ of even vertices of $\tilde{P}(v)$ contains $t$. Therefore $\tilde{P}(v)$ consists of three pieces, the first of which may be empty:

$$\tilde{P}(v) = (\tilde{P}(v) - \tilde{P}(w), \tilde{P}(w) - \tilde{P}(t), \tilde{P}(t)).$$

The proof that $\tilde{P}(v)$ is simple relies on the decomposition property (iv). The vertex $x$, which precedes $\bar{s}$ in $P(s)$, is an inner point $x \in I$. According to (iv), deletion of the edge $(\bar{s}, x)$ decomposes $A$ into two components $H_x$ and $H_r$ the latter of which contains the root $r$.

We note that $x \in I$ does not occur in $P(t)$. Otherwise, it would be odd in both $P(s)$ and $P(t)$, and lemma 1 would imply $x \in P(b)$. This contradicts $x \in P(s) - P(\bar{s}) \subseteq P(s) - P(b)$. Since $x \notin P(t)$, the path $P(t)$ cannot enter the component $H_x$ because by (iv) every path from $H_x$ to $r$ passes through $x$. Thus

(vii) $$P(t) \subseteq H_r.$$

Further we note that

(viii) $$P(s) - P(\bar{s}) \subseteq H_x.$$

Indeed, since $P(s)$ is simple, the edge $(\bar{s}, x)$ occurs exactly once in $P(s)$, bisecting this path. The piece $P(s) - P(\bar{s})$ lies in $H_x$, the piece $P(\bar{s})$ in $H_r$.

It follows from (vii) and (viii) that $\{P(s) - P(\bar{s})\} \cap P(t) = \phi$, and as a consequence, the predecessors of the even vertices of $P(t)$ are not altered. Hence

(ix) $$\tilde{P}(t) = P(t).$$

The third piece of $\tilde{P}(v)$ is therefore simple. The same holds for the two other pieces by virtue of (vi), and since the reverse of the path $\tilde{P}(w) - \tilde{P}(t)$ is part of the simple path $P(s) - P(\bar{s})$ by definition of $B$. Thus the proof of the lemma reduces to verifying

(x) $$\{\tilde{P}(w) - \tilde{P}(t)\} \cap P(t) = \phi$$

(xi) $$\{P(v) - P(w)\} \cap P(t) = \phi$$

(xii) $$\{P(v) - P(w)\} \cap \{\tilde{P}(w) - \tilde{P}(t)\} = \phi.$$

Now (x) follows from (vii) and (viii) since $\tilde{P}(w) - \tilde{P}(t) \subseteq P(s) - P(\bar{s})$. To prove (xi), we note that $w \in P(s) - P(\bar{s}) \subseteq H_x$. Hence $(\bar{s}, x)$ bisects $P(w)$. Since $P(v) \supseteq P(w)$, $(s, x)$ bisects $P(v)$ also, and we have

$$P(v) - P(w) \subseteq H_x.$$

This, together with (vii) proves (xi).

Finally assume $z \in \{P(v) - P(w)\} \cap \{\tilde{P}(w) - \tilde{P}(t)\}$. Without restriction of generality we may assume that $z$ is even in $P(v)$; otherwise we replace $z$ by $m(z)$. $z$ occurs in $P(s) - P(\bar{s})$ since $\tilde{P}(w) - \tilde{P}(t) \subseteq P(s) - P(\bar{s})$. If $z$ is even in $P(s)$, we have $w \leq s$ as a consequence of $w \leq z$ and $z \leq s$. But this contradicts the fact that $w$ is odd in $P(s)$. If, on the other hand, $z$ is odd in $P(s) - P(\bar{s})$, then $w$ would not be the first even vertex of this kind in $\tilde{P}(v)$. This proves (xii) and completes the proof of lemma 3.

## 5. Algorithms $\mathfrak{T}$ and $\mathfrak{M}$

We proceed to describe two algorithms which are based on successively enlarging labeled subgraphs by forward and blossom steps. The algorithm $\mathfrak{T}$ determines all vertices of a given matched graph $(G, M)$ that are legally accessible from a given exposed vertex $r$. It also arranges the connecting alternating paths in a treelike manner, thereby proving the main theorem. The algorithm $\mathfrak{M}$ is essentially based on the algorithm $\mathfrak{T}$, however its purpose is the construction of a maximum matching.

The algorithm $\mathfrak{T}$ goes as follows: Start out with the trivial labeled subgraph $L = (A, \Omega, I, p)$ whose graph $A$ consists of the root $r$ only. $r$ is labeled "outer", that is, $\Omega = \{r\}$ and $I := \phi$. The predecessor function need not be defined since its region of definition $\Omega - \{r\}$ is vacuous. Then enlarge $L$ successively by forward and blossom steps until a subgraph is reached that permits no further enlargement. We call such a labeled subgraph *terminal*. A terminal subgraph must be reached eventually since each step increases the number of edges in $A$, and this number is bounded by the number of edges in $G$. Terminal labeled subgraphs correspond to Hungarian trees (Edmonds [1]), if $M \cap A$ is part of a maximum matching.

LEMMA 4. *If* L = (A, Ω, I, p) *is terminal, then the vertices in* Ω *are precisely those vertices that are legally accessible from* r.

[5] This argument is based on the following simple lemma: *Let* C *be a closed path in a graph* G. *If* C *passes through the vertex* x *exactly once, then* C *contains a circuit through* x.

PROOF: Let $v \notin \Omega$ be a vertex which is joined to $r$ by a simple even alternating path $V$. It is no restriction of generality to assume that all other even vertices of $V$ belong to $\Omega$. Denote by $w$ the even vertex in $V \cap \Omega$ closest to $v$. Thus $(v, m(v), w)$ is a bi-edge. Now we treat two cases.

CASE 1: $v \notin I$. Then $v \notin \Omega \cup I$ and so $m(v) \notin \Omega \cup I$ (property (i)). Hence the bi-edge $(v, m(v), w)$ can be added to $L$ (forward step), contradicting the fact that $L$ is terminal.

CASE 2: $v \in I$. Hence $m(v) \in \Omega$ according to (ii), and deleting the edge $(v, pm(v))$ decomposes $A$ into two components $H_v$ and $H_r$ according to (iv). The edge of $V$ meeting $v$ is $(v, m(v))$, which leads the path $V$ into $H_v$. Since $V$ has no double points, it cannot contain the edge $(v, pm(v))$. Therefore, since $(v, pm(v))$ alone joins $H_v$ to the rest of $A$ and since $r$ is not in $H_v$, $V$ must leave $A$ eventually. Let $(z, u)$ with $z$ in $A$ be the first edge of $V$ after $v$ not in $A$. By virtue of property (i), the edge $(z, u)$ is not matched. Hence $z$ is odd, and $u$ is even in $V$. Since $u \neq v$, we have $u \in \Omega$ by hypothesis. This implies $z \in I$, because otherwise $z$ and $u$ would define a blossom containing $v$ showing $L$ to be not terminal.

The part of the path $V$ between $w$ and $z$, but in reverse order, forms together with $P(m(v))$ the path.

$$Z = (z, m(z), \ldots, w, P(m(v))) \subseteq A,$$

that joins $z$ to the root $r$. $Z$ does not contain the edge $(z, pm(z))$. Otherwise, $z$ would occur twice in $Z$, and this is impossible since $(z, m(z), \ldots, w)$ is simple as part of $V$, and since $z \in H_v$ and $P(m(v)) \cap H_v = \{m(v), v\}$. The existence of $Z$ contradicts the fact that, according to (iv), the edge $(z, pm(z))$ separates $z$ from $r$ in $A$. Thus there are no legally accessible vertices outside $\Omega$. That all vertices of $\Omega$ are indeed legally accessible follows by (v). This completes the proof of the lemma.

Algorithm $\mathfrak{T}$ proves our theorem for exposed roots $r$. However, as we already pointed out in section 2, this is sufficient to prove the theorem for general roots.

Algorithm $\mathfrak{M}$ starts out with any matched graph $(G, M)$. If there are no exposed vertices, then the matching is maximum. Otherwise choose an exposed vertex $e$ as a root, and employ algorithm $\mathfrak{T}$ to find outer vertices connected to $e$. If some outer vertex $v$ is adjacent to an exposed vertex $f$ different from $e$, then the edge $(f, v)$ together with the back-path $P(v)$ forms an augmenting path. Replacing matched edges by unmatched ones, and vice versa, along this path yields a new matching $M$ of higher cardinality. Algorithm $\mathfrak{T}$ then is repeated with one of the remaining exposed vertices as roots.

If algorithm $\mathfrak{T}$ does not produce an augmenting path, then we consider the terminal labeled subgraph $A$ and the graph $\bar{G} := G - A$ that results by deleting from $G$ all vertices of $A$ and all edges adjoining $A$. $\bar{G}$ is again a matched graph, and algorithm $\mathfrak{M}$ is applied to it. This procedure is justified by the following lemma due to Edmonds [1]:

LEMMA 5. *If* A *is a terminal labeled subgraph none of whose outer vertices are joined to an exposed vertex by an edge not in* A, *then the matching* M *of* G *is maximum if and only if its restriction to* G − A *is maximum.*

PROOF: The nontrivial direction is to show that if $M$ is not maximum, then neither is its restriction to $G - A$. If $e$ is the root of $A$, then $e$ is not the end-vertex of an augmenting path. Indeed, if there exists an augmenting path $P$ joining $e$ to an exposed vertex $f$, then the vertex $v$ which is next to $f$ in $P$ is legally accessible from $e$. Hence $v \in \Omega$ according to lemma 4. On the other hand, $f \notin A$, and therefore $(v, f)$ is not an edge of $A$, contradicting the hypothesis of the lemma.

Now let the augmenting path $V$ join the two exposed vertices $f$ and $g$, both in $G - A$. We have to show

(xiii) $$V \subseteq G - A.$$

Assume $V \cap A \neq \phi$. Since $x \in V \cap A$ implies $m(x) \in V \cap A$, we conclude $V \cap \Omega \neq \phi$. Let $v \in V \cap \Omega$, then there exists a back-path $P(v)$ connecting $v$ and $e$. Let $w$ be the vertex in $P(v) \cap V$ which is closest to $e$ in $P(v)$. $w$ is odd in $P(v)$; otherwise, $m(w) \in P(v) \cap V$ would be closer to $e$ in $P(v)$. $w$ divides the path $V$ into two pieces; the piece which contains $m(w)$, forms together with $P(m(w))$ an augmenting path ending at $e$. But such an augmenting path has been shown not to exist. This proves (xiii), and thereby the lemma.

## 6. An Example by E. Johnson

Berge [4] considers a generalization of the maximum matching problem. Given for each vertex $v$ of a graph $G$ a nonnegative integer $d(v)$, one calls a *degreeconstrained subgraph* a subgraph $M \subseteq G$ whose degrees at each vertex $v$ are bounded above by $d(v)$. The problem is to find a *degreeconstrained* subgraph with a maximum number of edges. The maximum matching problem results if $d(v) = 1$ for all vertices $v$ of $G$. Maximum degreeconstrained subgraphs are again characterized by the absence of "augmenting paths". Augmenting paths are again alternating paths without multiple edges; multiple points however are permitted (Berge [4], compare also Goldman [6]).

Recently, Edmonds [3] extended his algorithm to an efficient solution algorithm for the above and even more general problems. This leads to the question of extending the predecessor technique described in this paper.

Alternating paths and bi-edge graphs can be defined for every pair $(G, S)$, where $S$ is any set of edges of $G$, not necessarily a matching. The theory of degree-constrained subgraphs suggests defining *legal* alternating paths as alternating paths without multiple edges, but possibly multiple vertices. The predecessor technique now requires that our theorem holds for general $S$ with the above concept of legality.

This conjecture is refuted by the following example due to E. Johnson. The graph $G$ of this example is presented in figure 7; heavy lines distinguish the edges in $S$.

Choosing vertex *1* as the root, all vertices are legally accessible by alternating paths of even length. From

vertex *2*, for instance, the root can be reached by the paths (*2, 3, 5, 4, 3, 6, 7, 8, 9, 10, 8, 2, 1*) or (*2, 8, 10, 9, 8, 7, 6, 3, 4, 5, 3, 2, 1*). Up to interchanges of *4* with *5* and *9* with *10*, these are the only simple alternating paths joining *2* to the root. If the first path is to appear in a tree $T \subseteq B(G, S)$, then the predecessor of *8* must be *6*. But up to an interchange of *9* with *10*, the only back-path of *7* is (*7, 8, 10, 9, 8, 2, 1*) requiring *1* to be the predecessor of *8*. A similar contradiction appears if the second back-path of *2* is chosen to be in the tree. Thus the legal back-paths in figure 7 cannot be arranged in a treelike manner. Without the legality requirement, however, no difficulty would have arisen.

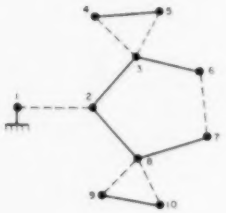



FIGURE 7

This example proves that the maximum degree-constrained subgraph problem cannot be solved by a direct extension of the predecessor algorithm. Another generalization of the maximum matching problem, also solved by Edmonds [2], consists in maximizing a weighted sum of the edges. The shrinking technique of Edmonds appears to be more natural for this problem also.

## 7. References


[1] Edmonds, J., Paths, trees, and flowers (to appear, Canadian J. Math. (May 1965).

[2] Edmonds, J., Maximum matching and a polyhedron with 0, 1–vertices, J. Res. NBS **69B** (Math. and Math. Phys.) Nos. 1 and 2, 125–130 (Jan.–June 1965).

[3] Edmonds, J., Optimum degreeconstrained subgraphs (in preparation).

[4] Berge, C., The Theory of Graphs and Its Applications (English translation from French), John Wiley & Sons, Inc., New York, N.Y.).

[5] Edmonds, J., Covers and Packings in a Family of Sets, Bull. Amer. Math. Soc. **68,** No. 5, (Sept. 1962).

[6] Goldman, A. J., Optimal matchings and degreeconstrained subgraphs, J. Res. NBS **68B** (Math. and Math. Phys.), No. 1, 27–29 (1964).


(Paper 69B1–139)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# On Measurable Sets and Functions

A. J. Goldman

(October 6, 1964)

The relation $L=f^{-1}(B)$, where $L$ and $f$ are Lebesgue measurable and $B$ is a Borel set, is studied. Any one of $L$, $B$, $f$ can be specified and the relation is solvable; one can also specify two of the three possible pairs. The relation characterizes (in a sense made precise in the text) the classes of Lebesgue measurable functions and sets; that it does so for the class of Borel sets as well is left as a conjecture, whose truth would imply that the functions which preserve Lebesgue measurability as second composition factors [i.e., $g$ in $g(h(x))$] are precisely the Borel measurable functions.

The questions treated in this note, though mainly elementary, occur so naturally in connection with the basic concepts of measure and integration theory as to warrant unified presentation. For concreteness we deal exclusively with subsets of the real line $R$, and with real-valued functions defined on $R$. The symbols $(BS)$, $(ZS)$, and $(LS)$ denote the respective classes of Borel sets, sets of zero measure, and Lebesgue-measurable sets; we recall that $L\epsilon(LS)$ if and only if $L$ has a representation of the form

$$L=(B-Z)\cup(Z-B) \qquad B\epsilon(BS),\ Z\epsilon(ZS). \tag{1}$$

The symbols $(BF)$ and $(LF)$ denote the respective classes of Borel-measurable functions and Lebesgue-measurable functions; we recall that $f\epsilon(LF)$ if and only if

$$L=f^{-1}(B)\epsilon(LS) \qquad \text{for all } B\epsilon(BS). \tag{2}$$

Our first theme is the *solvability* of relation (2) when various subsets of its "variables" $(L, B, f)$ are specified. For example, given $B\epsilon(BS)$ we can trivially find $L\epsilon(LS)$ and $f\epsilon(LF)$ to satisfy (2) by choosing $L=B$ and $f=$(identity); given $f\epsilon(LF)$ we can trivially find $B\epsilon(BS)$ and $L\epsilon(LS)$ to satisfy (2) by choosing *any* $B\epsilon(BS)$ and setting $L=f^{-1}(B)$. A third case (in which $L$ is specified) is treated in the following theorem.

THEOREM 1: *For any* $L\epsilon(LS)$, *there is a* $B\epsilon(BS)$ *differing from* $L$ *by at most a set of measure zero, and an* $f\epsilon(LF)$ *differing from the identity on at most a set of measure zero, such that* $L=f^{-1}(B)$.

PROOF: If $L=\phi$ or $L=R$, take $B=L$ and $f=$(identity). If $L\epsilon(ZS)$ but $L\neq\phi$, there exists $B\subset L$ with $B\epsilon(BS)$ and $B\neq\phi$; let $f$ be the identity on $(R-L)\cup B$ and [1] let $f(L-B)=\{x\}$ for some $x\epsilon B$. If $R-L\epsilon(ZS)$ and $R-L\neq\phi$, there exists $B'\subset R-L$ with $B'\epsilon(BS)$ and $B'\neq\phi$; let $f$ be the identity on $L\cup B'$, and [2] let $f(R-L-B')=\{x\}$ for some $x\epsilon B'$ (here $B=R-B'$).

Finally, suppose none of the above situations holds. Consider a representation (1) of $L$, and let $f$ be the identity on $R-Z$. Since it is not true that $B=\phi$, we can define $f$ on $Z-B$ so that $f(Z-B)\subset B$. Since it is not true that $R-B=\phi$, we can define $f$ on $Z\cap B$ so that $f(Z\cap B)\subset R-B$. This completes the proof.

It is natural next to consider the solvability of (2) when *two* of $(L, B, f)$ are specified. Given $B\epsilon(BS)$ and $f\epsilon(LF)$, (2) serves to define an $L\epsilon(LS)$ which satisfies (2). The case in which $B$ and $L$ form the specified pair is treated in the following theorem.

THEOREM 2: *For any* $B\epsilon(BS)$ *and* $L\epsilon(LS)$, *with sole exceptions* $(B=\phi,\ L\neq\phi)$ *and* $(B=R,\ L=\phi)$, *there is an* $f\epsilon(LF)$ *such that* $L=f^{-1}(B)$.

PROOF: First suppose $B=\phi$; then if $L=\phi$ any $f\epsilon(LF)$ will do, while if $L\neq\phi$ no $f$ will do. Next suppose $B=R$; if $L\neq\phi$ we can choose $f\epsilon(LF)$ so that $f(R)\subset L$, while if $L=\phi$ no $f$ will do. Finally, if $B\neq\phi$ and $B\neq R$, then we can define $f$ on $L$ so that $f(L)\subset B$, and on $R-L$ so that $f(R-L)\subset R-B$.

The remaining case is that in which $L\epsilon(LS)$ and $f\epsilon(LF)$ are specified. One cannot always find $B\epsilon(BS)$ to satisfy (2) (suppose e.g., that $L\epsilon(LS)-(BS)$ and $f=$(identity)), so that the question must be modified. One might ask for which $f\epsilon(LF)$ it is true that to *each* $L\epsilon(LS)$ there corresponds a $B\epsilon(BS)$ obeying (2); the answer is "no $f$" even without the measurablility requirement on $f$, since the cardinality $2^c$ of $(LS)$ exceeds the cardinality $c$ of $(BS)$. A second modified version is also uninteresting, as the next theorem shows:

THEOREM 3: $L=\phi$ *and* $L=R$ *are the only* $L\epsilon(LS)$ *such that for each* $f\epsilon(LF)$, *there exists* $B\epsilon(BS)$ *satisfying relation* (2).

PROOF: $L=\phi$ and $L=R$ *are* solutions, since one can take $B=L$ independent of $f$. For any other $L\epsilon(LS)$, choose $f$ to be the characteristic function of some $L'\epsilon(LS)$ different from both $L$ and $R-L$; then $f^{-1}(B)\neq L$ for *all* $B\subset R$, since $f^{-1}(B)$ will be one of the four sets $\phi$, $R$, $L'$, $R-L'$ according to the membership or nonmembership of 0 and 1 in $B$.

Our second theme concerns the role of relation (2) in *characterizing* the three classes involved ($(LF)$,

[1] Omit this clause if $B=L$.
[2] Omit this clause if $B'=R-L$.

$(LS)$, $(BS)$), in the sense of the equations

$$(LF)=\{f:(f:R\to R), f^{-1}(B)\epsilon(LS) \text{ for all } B\epsilon(BS)\},$$
$$(LS)=\{L:L\subset R, L=f^{-1}(B) \text{ for some } B\epsilon(BS) \text{ and } f\epsilon(LF)\},$$
$$(BS)=\{B:B\subset R, f^{-1}(B)\epsilon(LS) \text{ for all } f\epsilon(LF)\}.$$

The first of these equations holds trivially; i.e., the relation (2) used to define $(LF)$ certainly characterizes $(LF)$. Theorem 1 shows that the right side of the second equation contains $(LS)$; since the inclusion in the opposite direction is trivial, $(LS)$ is also characterized by (2). The right side of the third equation clearly contains $(BS)$, so that what remains to be proved is equivalent to the following statement, which the writer has been unable to settle:

CONJECTURE: *If* S *is not a Borel set then for at least one* $f\epsilon(LF)$, $f^{-1}(S)$ *fails to be Lebesgue measurable.*

Our final theme is the preservation of measurability under function composition. Composition will be denoted by an asterisk, i.e., $(f*g)(x)=f(g(x))$. We set

$$(LCF)=\{f:(f:R\to R), f*g\epsilon(LF) \text{ for all } g\epsilon(LF)\}$$

where "$LCF$" is a mnemonic for "left composition factor." Taking $g$ as the identity shows that $(LCF)\subset(LF)$; it is an unpleasant fact that the inclusion is strict. Some texts include a proof that $(LCF)$ contains the continuous functions, while others give the sharper result that the class of Borel measurable functions $(BF)\subset(LCF)$. We shall show why this may be the best possible result:

THEOREM 4: (BF)=(LCF) *if the Conjecture is true.*

PROOF: First assume $f\epsilon(BF)$; then for any $g\epsilon(LF)$ and $B\epsilon(BS)$ we have $f^{-1}(B)\epsilon(BS)$ and therefore

$$\{x:f(g(x))\epsilon B\}=g^{-1}(f^{-1}(B))\epsilon(LS)$$

so that $(f*g)^{-1}(B)\epsilon(LS)$ for all $B\epsilon(BS)$. Thus $f*g\epsilon(LF)$ for all $g\epsilon(LF)$, proving $f\epsilon(LCF)$. Next assume $f\epsilon(LF)-(BF)$. Then $f^{-1}(B)\epsilon(LS)-(BS)$ for some $B\epsilon(BS)$. By the Conjecture, there exists $g\epsilon(LF)$ for which $g^{-1}(f^{-1}(B))=(f*g)^{-1}(B)$ is *not* in $(LS)$; thus $f*g$ is not in $(LF)$, and hence $f$ is not in $(LCF)$, completing the proof.

Similarly, we define

$$(RCF)=\{g:(g:R\to R), f*g\epsilon(LF) \text{ for all } f\epsilon(LF)\}.$$

Taking $f$ as the identity shows that $(RCF)\subset(LF)$, and it is known [3] that $(RCF)$ does not even contain all continuous strictly monotone functions. For an alternate characterization of $(RCF)$, we set

$$(SLF)=\{g:(g:R\to R), g^{-1}(L)\epsilon(LS) \text{ for all } L\epsilon(LS)\}$$

where "$SLF$" is a mnemonic for "strongly Lebesgue measurable function."

THEOREM 5: (SLF)=(RCF)

PROOF: First assume $g\epsilon(SLF)$; then for any $f\epsilon(LF)$ and $B\epsilon(BS)$ we have $f^{-1}(B)\epsilon(LS)$ and therefore

$$(f*g)^{-1}(B)=g^{-1}(f^{-1}(B))\epsilon(LS).$$

Therefore $f*g\epsilon(LF)$ for each $f\epsilon(LF)$; i.e., $g\epsilon(RCF)$. Next assume $g\epsilon(LF)-(SLF)$; then there exists $L\epsilon(LS)$ for which $g^{-1}(L)$ is not in $(LS)$, and by Theorem 1 $L=f^{-1}(B)$ for some $f\epsilon(LF)$ and $B\epsilon(BS)$. Thus $(f*g)^{-1}(B)$ is not in $(LS)$, and so $g\epsilon(LF)-(RCF)$, completing the proof.

It would be interesting to explore the class $(SLF)$ more thoroughly.

---

[3] See p. 83 of Halmos' "Measure Theory," van Nostrand, 1950.

(Paper 69B1–140)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Characteristic Relations for Nonperiodic Solutions of Mathieu's Equation*

T. Tamir and H. C. Wang

Electrophysics Department, Polytechnic Institute of Brooklyn, Brooklyn, N.Y.

(October 30, 1964)

A simple compact formulation is presented for the eigenvalue $\lambda$ of the Mathieu differential equation $y''+(\lambda-2h^2\cos 2\theta)y=0$. The relations obtained thereby are employed to examine certain characteristic features of the solutions $y=e^{i\nu\theta}\sum_n c_n e^{2in\theta}$ in both stable ($\nu$ real) and unstable ($\nu$ complex) regions.

In the vicinity of the boundaries ($\nu=m=$ integer) which separate these regions, the expressions derived for $\lambda$ are nonsingular and continuous; at the boundaries, these expressions are identical to available expansions which become therefore special cases of a more general formulation. The results are used to determine the magnitudes of the harmonic coefficients $c_n$ and the instability intervals. The latter are shown to be intimately related to restrictions on the complex value of the characteristic exponent $\nu$. In addition, simple approximations that are suitable for perturbation problems ($h^2$ small) are derived for all of the characteristic quantities.

## 1. Introduction

Hill's differential equation and its special case, the Mathieu equation with real coefficients

$$\frac{d^2y}{d\theta^2}+(\lambda-2h^2\cos 2\theta)y=0 \tag{1.1}$$

were first examined in connection with the vibration of elliptic membranes [1][1] and the motion of celestial bodies [2]. More recent studies were motivated by applications in the fields of acoustical and electromagnetic diffraction, frequency modulation, electromechanical transducer design, elliptic waveguides, etc.; since it is impractical to provide a complete bibliography within the present framework, the reader is referred to several excellent textbooks [3–6], which contain lengthy lists of references.

Most of the investigations of the Mathieu equation (1.1) were concerned primarily with only a very limited range of solutions—the periodic ones. Thus, if the solution to (1.1) is expressed in the Floquet form

$$y(\theta)=e^{\pm i\nu\theta}\sum_{n=-\infty}^{\infty}c_n(\pm i\nu)e^{2in\theta} \tag{1.2}$$

where the $\pm$ signs relate to the two possible characteristic solutions, these will be periodic whenever the characteristic exponent $\nu$ is an integer ($\nu=m=0, 1, 2, 3, \ldots$). In those restricted cases, the solutions $y(\theta)$ were extensively tabulated [7].

Of a somewhat different type are problems which do not require periodic solutions; in those cases, $\nu$ is not necessarily a real integer and it may also become complex. Thus, the Mathieu equation (1.1) may express the propagation of electromagnetic waves in a medium that possesses dielectric or magnetic properties which vary periodically [8, 9, 10]; in that case, $\theta$ represents a

*This paper is based on a portion of a dissertation submitted by H. C. Wang in partial fulfillment of the requirements for the Ph. D. (Electrophysics) degree at the Polytechnic Institute of Brooklyn, New York.

[1] Figures in brackets indicate the literature references at the end of this paper.

space variable and $c_n(\pm i\nu)$ refer to the amplitudes of the space harmonics. An analogous situation occurs in the case of certain types of oscillators [11] (particularly those involving parametric action) for which $\theta$ is interpreted as a time variable whereas $c_n(\pm i\nu)$ relate to the amplitudes of the temporal harmonics. In all of these cases, tabulated results are few and restricted in range [6, 12].

The significance of the characteristic exponent $\nu$ can usually be related directly to the wave character of the physical phenomena described by the Mathieu equation (1.1). Hence the determination of $\nu$, i.e., its relation in terms of $\lambda$ and $h^2$ or, alternatively, the determination of $\lambda=\lambda(\nu, h^2)$, are usually of considerable importance. It is also recognized that, when a forcing term is present, the source (generator) element usually prescribes the wavelike nature of the phenomena as reflected through $\nu$; hence $\lambda$, rather than $\nu$, represents the unknown physical quantity to be determined. One may then employ expressions that are found in the literature [1–7] and which relate $\lambda$ in terms of power series of $h^2$ for integral values of $\nu$. For $\nu$ not an integer, the power series

$$\lambda=\nu^2+\frac{1}{2(\nu^2-1)}h^4+\frac{5\nu^2+7}{32(\nu^2-1)^3(\nu^2-4)}h^8+\ldots \tag{1.3}$$

is available; however, this relation converges poorly for $\nu$ close to an integer and its applicability is therefore limited.

The present paper presents a method for deriving explicit expressions for $\lambda=\lambda(\nu, h^2)$ which hold for arbitrary values of $\nu$. The expansions involved are based on the continued fraction relations [3–7] of the Mathieu equation and is an extension of a method for obtaining power series for $\nu$ integer [13]. The expressions obtained herein are not singular for $\nu$ integer, they are applicable for both real and complex values of $\nu$ and appear to converge rapidly. A different method which also converges rapidly was recently proposed by Schäfke, Ebert, and Groh [14]; in contrast to their results, the expressions presented here are explicit and they are particularly simple when approximate expressions are required for small $h^2$.

In addition to expansions for the eigenvalue $\lambda$, other characteristic parameters can be found via the expansions given here. Thus, the extent of the range for complex values of $\nu$ is discussed, as well as the relation between all of the harmonic amplitudes $c_n(\pm i\nu)$. It is also shown that, in the limit of $\nu$ becoming an integer, the results presented here are identical to those already available [1–7] for the Mathieu functions and it is suggested that the present method leads to a more systematic approach for future computations of the pertinent series.

Certain physical interpretations, notation and the representation of the various possible solutions are discussed in connection with the so-called "stability chart" in section 2. The derivation of the expansions for $\lambda=\lambda(\nu, h^2)$ is presented in section 3 and the pertinent formulas for small $h^2$ are given in section 4. Simplified expansions for $\nu$ integer are discussed in section 5 and the range for $\nu$ complex is evaluated in section 6. The magnitudes of the harmonics $c_n(\pm i\nu)$ are dealt with in section 7. Finally, some comments are made in section 8 concerning the range for $h^2$ wherein the relations presented here are expected to be valid and to converge rapidly.

## 2. Stability Chart

The harmonic amplitudes $c_n=c_n(\pm i\nu)$ are interrelated via the infinite continued fraction expansions [3–6]:

$$\frac{c_n}{c_{n-1}}=-\frac{h^2}{L_n-}\frac{h^4}{L_{n+1}-}\frac{h^4}{L_{n+2}-}\cdots\cdots \tag{2.1}$$

$$\frac{c_n}{c_{n+1}}=-\frac{h^2}{L_n-}\frac{h^4}{L_{n-1}-}\frac{h^4}{L_{n-2}-}\cdots\cdots \tag{2.2}$$

where

$$L_n=(\nu+2n)^2-\lambda. \tag{2.3}$$

By combining eqs (2.1) and (2.2), one obtains

$$\frac{L_n}{h^4}=\frac{1}{L_{n+1}}-\frac{h^4}{L_{n+2}}-\cdots\cdots+\frac{1}{L_{n-1}}-\frac{h^4}{L_{n-2}}-\cdots\cdots \tag{2.4}$$

which is essentially an equation for $\lambda=\lambda(\nu, h^2)$. By solving it, one may obtain the eigenvalue $\lambda$ and thence the harmonic amplitudes $c_n$ via (2.1) or (2.2).

Relation (2.4) is often presented [3–9] in the form of a so-called "stability chart", as shown in figure 1. This diagram illustrates some of the properties associated with the eigenvalue $\lambda$ and the characteristic exponent $\nu$. Thus, for $h^2$ and $\lambda$ real, $\nu$ may be either real or complex. In the latter case:

$$\nu=m+i\alpha \tag{2.5}$$

where $\alpha$ is real and $m=0, 1, 2, 3\ldots$. For simplicity, the real part of $\nu$ will always be taken as positive, in both the real and the complex $\nu$ cases.

Regions with real and complex values of $\nu$ are called "stable" and "unstable" regions (or bands), respectively, in the literature [3–12]. A systematic definition of this terminology can be traced to van der Pol and Strutt [15]: the justification is afforded by the fact that, if $\nu$ is complex, the characteristic solutions in eq (1.2) are not bounded as $|\theta|\to\infty$. It is recognized that the "unstable" solutions should not necessarily be neglected in actual physical situations. Thus, when $\theta$ is bounded, both complex values of $\pm\nu$ are admissible; also, in certain unbounded situations, one of the solutions may be retained and only the other need be rejected. The latter case is illustrated by the radiation condition in electromagnetic fields [10].

From figure 1, one sees that unstable regions (bands) alternate with stable ones. The lines $\nu=m$ consist of two separate branches (for all $m\neq 0$) which join on the $\lambda$ axis and contain a region with $\nu$ complex ($=m+i\alpha$); it may be shown that the two $\nu=m$ lines touch each other on the $\lambda$-axis with an order of tangency of $m-1$. As a consequence, for a fixed $h^2$, the width of the unstable band decreases as $m$ increases.

As indicated in figure 1, the curves for $\nu=$const. exhibit different variations in the stable and unstable regions. Those with $\nu$ real start from the $\lambda$-axis (at $\nu^2=\lambda$) and proceed towards increasing values of $h^2$; on the other hand, the curves for $\nu$ complex exhibit a minimum with respect to $h^2$ and are double-valued with respect to $\lambda$.

Whereas the stability diagram may be used directly [9, 10] to obtain one of the parameters $\lambda$, $\nu$, or $h^2$ when the other two are given, this procedure usually leads to a graphical construction and therefore its accuracy and convenience are somewhat limited. Instead, simple analytical expressions for $\lambda=\lambda(\nu, h^2)$ may be used and their derivation is given in the following section.

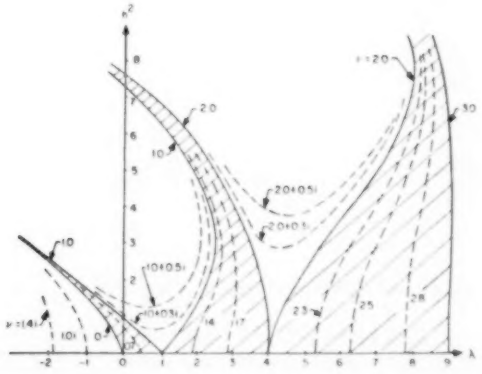



FIGURE 1. *The Mathieu stability diagram.*

## 3. Relations for the Eigenvalue λ

In the present section, relations in the form $\lambda=\lambda(\nu,\ h^2)$ are obtained via the continued fractions of eq (2.4). When properly expanded as power series, these relations lead to equations for the eigenvalue λ which, unlike eq (1.3), are not singular at the band edges ($\nu=m=0,\ 1,\ 2,\ 3\ .\ .\ .$). In order to better appreciate the procedure employed here, it is pertinent to illustrate first the derivation of eq (1.3).

By setting $n=0$ in eq (2.4), one obtains

$$L_0=\nu^2-\lambda=\frac{h^4\rfloor}{L_1}-\frac{h^4\rfloor}{\lfloor L_2}-\frac{h^4\rfloor}{\lfloor L_3}-\cdots+\frac{h^4\rfloor}{L_{-1}}-\frac{h^4\rfloor}{\lfloor L_{-2}}-\frac{h^4\rfloor}{\lfloor L_{-3}}\cdots \tag{3.1}$$

which may be written as

$$\lambda=\nu^2-h^4(K^+ + K^-) \tag{3.2}$$

where

$$K^{\pm}=\frac{1\rfloor}{L_{\pm 1}}-\frac{h^4\rfloor}{\lfloor L_{\pm 2}}-\frac{h^4\rfloor}{\lfloor L_{\pm 3}}-\cdots \tag{3.3}$$

Equation (3.2) is now viewed as an iteration formula

$$\lambda_{i+1}=\nu^2-h^4[K^+(\lambda_i)+K^-(\lambda_i)] \tag{3.4}$$

for finding $\lambda_{i+1}$ when a trial value $\lambda_i (i=0,\ 1,\ 2,\ 3\ .\ .\ .)$ is assumed. To obtain an analytical result in terms of a power series in $h^2$, it is necessary to expand $K^{\pm}$ as a power series. This is achieved via a useful series expansion for continued fractions [16] which, in the present case, is written as

$$K^{\pm}=\frac{1}{L_{\pm 1}}+\frac{h^4}{L_{\pm 1}^2\left(L_{\pm 2}-\dfrac{h^4}{L_{\pm 1}}\right)}+\frac{h^8}{L_{\pm 1}^2\left(L_{\pm 2}-\dfrac{h^4}{L_{\pm 1}}\right)^2\left(L_{\pm 3}-\dfrac{h^4\rfloor}{L_{\pm 2}}-\dfrac{h^4}{\lfloor L_{\pm 1}}\right)}$$

$$+\frac{h^{12}}{L_{\pm 1}^2\left(L_{\pm 2}-\dfrac{h^4}{L_{\pm 1}}\right)^2\left(L_{\pm 3}-\dfrac{h^4\rfloor}{L_{\pm 2}}-\dfrac{h^4}{\lfloor L_{\pm 1}}\right)^2\left(L_{\pm 4}-\dfrac{h^4\rfloor}{L_{\pm 3}}-\dfrac{h^4\rfloor}{\lfloor L_{\pm 2}}-\dfrac{h^4}{\lfloor L_{\pm 1}}\right)}+\cdots \tag{3.5}$$

This expansion may be cast in the more compact form

$$K^{\pm}=\sum_{k=0}^{\infty} B_{\pm k}B_{\pm(k+1)}h^{4k} \tag{3.6}$$

where

$$B_0=1;\ B_{\pm k}=\prod_{j=1}^{k} A_{\pm j} \tag{3.7}$$

and

$$A_{\pm j}=\frac{1\rfloor}{L_{\pm j}}-\frac{h^4\rfloor}{\lfloor L_{\pm(j-1)}}-\frac{h^4\rfloor}{\lfloor L_{\pm(j-2)}}-\cdots-\frac{h^4\rfloor}{\lfloor L_{\pm 2}}-\frac{h^4}{\lfloor L_{\pm 1}} \tag{3.8}$$

$$(j=1,\ 2,\ 3\ .\ .\ .;\ k=0,\ 1,\ 2,\ .\ .\ .)$$

and it is understood that $A_{\pm 1}=\dfrac{1}{L_{\pm 1}}$.

It is recognized that the expansion in eqs (3.5) or (3.6) would be a power series in terms of $h^4$ if terms with $h^4$ did not also appear in $A_{\pm j}$. However, if $h^4$ is sufficiently small, one may employ the binomial series expansion to convert the above in a power series. Thus:

$$K^{+}=\frac{1}{L_1}+\frac{h^4}{L_1^2L_2}\left(1-\frac{h^4}{L_1L_2}\right)^{-1}+\frac{h^8}{L_1^2L_2^2L_3}\left(1-\frac{h^4}{L_1L_2}\right)^{-2}\left[1-\frac{h^4}{L_2L_3}\left(1-\frac{h^4}{L_1L_2}\right)^{-1}\right]^{-1}+\cdots$$

$$=\frac{1}{L_1}+\frac{h^4}{L_1^2L_2}+\left(1+\frac{L_3}{L_1}\right)\frac{h^8}{L_1^2L_2^2L_3}+\left(1+\frac{L_4}{L_2}+2\frac{L_3L_4}{L_1L_2}+\frac{L_3^2L_4}{L_1^2L_2}\right)\frac{h^{12}}{L_1^2L_2^2L_3^2L_4}+0(h^{16}) \tag{3.9}$$

and a similar power series for $K^{-}$. Introducing these series into eq (3.4), one obtains the required iteration formula which yields $\lambda$ in terms of $\nu$ and a power series in $h^4$. The proper starting value for the iteration process is obviously $\lambda_0=\nu^2$ for the stable regions and $\lambda_0=m^2$ ($m=0, 1, 2\ .\ .\ .$) for the unstable regions, since these are the values of $\lambda$ at $h=0$ (see fig. 1). It is observed that by taking these starting values of $\lambda_0$, the iteration formula (3.4) is in a form such that $\lambda_i$ is obtained to $0(h^{4i})$. Hence, if $\lambda$ is to be found to $0[h^{4(i+1)}]$, it is necessary to find the series expansions for $K^{\pm}$ only up to $0(h^{4i})$.

By using these considerations, one may verify that eq (3.2) yields formula (1.3). As already noted, however, this formula yields good results only for values of $\nu$ which are not too close to a nonzero integer. When this condition is satisfied and if $h^2$ is not too large, both the iteration (3.4) and eq (1.3) converge rapidly. This fast convergence is due to the fact that $\lambda_{i+1}$ in eq (3.4) was isolated out from the term $L_0$, as indicated in eq (3.1) and, of all $L_n$ terms, $L_0$ varies most rapidly with both $h^2$ and $\nu$, provided $0<$ real part of $\nu<1$. This property is verified by noting that, for small $h^2$, one may write $\lambda\cong\nu^2$ to yield

$$L_n\cong(\nu+2n)^2-\nu^2=4n(\nu+n) \tag{3.10}$$

which is most sensitive for $n=0$ and $n=-\nu=-m$, where $m$ is the integer that is closest to the real part of $\nu$. Hence, as long as $\nu$ is real and sufficiently smaller than unity, $L_0$ is the only sensitive term. However, in general *both* $L_0$ and $L_{-m}$ may become critical; hence, an iteration formulation such as eq (3.4) would have to be based on a value for $\lambda_{i+1}$ which is isolated out from *both* $L_0$ and $L_{-m}$. This is accomplished herein by employing certain properties of the pertinent functions $L_n$, as discussed below.

Consider first the problem of isolating the term $L_0$ out of a continued fraction such as eqs (2.1) or (2.2). If $|n|$ is large, $L_0$ will appear far down in the expansion and the process of finding its value will become quite tedious. It is therefore convenient to extend the concept of $K^{+}$ of eq (3.3) and define

$$K_n=\frac{1}{L_n}\Big\rfloor-\frac{h^4}{L_{n+1}}\Big\rfloor-\frac{h^4}{L_{n+2}}\Big\rfloor-\cdots\cdots \tag{3.11}$$

where the subscripts of the denominators form an *increasing* sequence. The term $L_0$ will appear only if $n\leq 0$; one may then write (see appendix):

$$K_n=A_n+\frac{B_n^2h^{-4n}}{S-C_nh^4}\qquad(n\leq 0) \tag{3.12}$$

where the terms $A_n$ and $B_n$ are given in eqs (3.8) and (3.7), respectively, with

$$S=L_0-K_1h^4 \tag{3.13}$$

$$A_0=C_0=0;\ C_n=\sum_{k=0}^{-(n+1)}B_{-k}B_{-(k+1)}h^{4k}\ (n<0) \tag{3.14}$$

and it is noted that the terms $A_n=A_n(\nu)$, $B_n=B_n(\nu)$ and $C_n=C_n(\nu)$ were defined in such a manner that, if expanded as power series in $h^4$, their expansion will start with a free term.

In addition to the simplification realized by writing $K_n$ in the form of eq (3.12), the terms $L_n$ possess a certain symmetry which is basic in the derivations obtained here. This feature is exhibited by taking $\nu$ as

$$\nu = m + \Delta \tag{3.15}$$

where

$$\Delta = \begin{cases} \beta \text{ (for } \nu \text{ real)} \\ i\alpha \text{ (for } \nu \text{ complex).} \end{cases} \tag{3.16}$$

For complex values of $\nu$, eqs (3.15) and (3.16) correspond to eq (2.5) and therefore $m$ denotes the pertinent unstable region. In the stable regions, $m$ is chosen so that $|\beta| \leq \frac{1}{2}$; hence $m$ is then the integer closest to $\nu$ in an arithmetic sense.

Defining the "complementary" value of $\nu$ as

$$\bar{\nu} = m - \Delta = 2m - \nu, \tag{3.17}$$

one recognizes that $\bar{\nu}$ corresponds to the complex conjugate value of $\nu$ if the latter is complex. Consider now any function $F = F(\nu)$; then

$$\bar{F} = \bar{F}(\nu) \triangleq F(\bar{\nu}), \tag{3.18}$$

where $\triangleq$ indicates equality by definition. If $F(\nu)$ is a polynomial, one may separate it as

$$F = Fe + Fo \tag{3.19}$$

where $Fe = Fe(m, \Delta)$ and $Fo = Fo(m, \Delta)$ are the functions containing only even or odd powers of both $m$ and $\Delta$. For complex values of $\nu$, $Fe$, and $Fo$ refer to the real and imaginary parts of $F$, respectively. One then also has

$$\bar{F} = Fe - Fo \tag{3.20}$$

$$\tilde{F}^2 \triangleq F\bar{F} = (Fe)^2 - (Fo)^2 \tag{3.21}$$

and it is recognized that, if $\nu$ is complex, the modulus $\tilde{F}$ is identical to the absolute value $|F|$. Applying definitions (3.17) and (3.18) to $L_n$ in eq (2.3), one gets

$$L_n = \bar{L}_{-(m+n)} \tag{3.22}$$

which is the "symmetry" feature alluded above. Its importance lies in the fact that, whereas only the first fraction (with increasing indices in $L_n$) of eq (2.4) is expressed via $K_n$ of eqs (3.11) and (3.12), the second (decreasing) continued fraction is expressible in terms of $\bar{K}_n$ via eq (3.22).

A crucial point in the above definitions is the fact that $\lambda$ is regarded as an independent parameter when taking the complementary of $F(\nu)$. In particular, one notes that $L_n = L_n(\nu, \lambda)$ which may also be written as $L_n = L_n[\nu, \lambda(\nu)]$; definition (3.18), however implies that $\bar{L}_n = L_n(\bar{\nu}, \lambda) = L_n[\bar{\nu}, \lambda(\nu)]$ and therefore $\bar{L}_n \neq L_n[\bar{\nu}, \lambda(\bar{\nu})]$.

The following discussion is more convenient if even and odd values of $m$ are viewed separately.

(a) $m$ odd: $m=2r-1$ $(r=1,2,3\ \ldots)$

For odd values of $m$, eq (2.4) may be written, with $n=1-r$, as

$$L_{1-r}-\cfrac{h^4}{L_{2-r}-\cfrac{h^4}{L_{3-r}-\cdots}}=\cfrac{h^4}{L_{-r}-\cfrac{h^4}{L_{-1-r}-\cfrac{h^4}{L_{-2-r}-\cdots}}} \tag{3.23}$$

which, by using relations (3.11), (3.18), and (3.22) becomes

$$\frac{1}{K_{1-r}}=\overline{K}_{1-r}h^4. \tag{3.24}$$

The sensitive $L_{-m}$ term is thus replaced with $\overline{L}_0$. One recalls that, for an iteration formula similar to that of eq (3.4), the term $\lambda_{i+1}$ needs to be isolated from the $L_0$ term, i.e., out of $S$ of eq (3.13) in the present case; in particular, since now $m\neq 0$, $\lambda_{i+1}$ will be obtained from the *even* part of $L_0$. Writing

$$X=S-C_{1-r}h^4 \tag{3.25}$$

the required iterated eigenvalue is given, via eqs (2.3), (3.13), and (3.25), by

$$\lambda^{(2r-1)}=m^2+\Delta^2-Xe-(Ke_1+Ce_{1-r})h^4 \tag{3.26}$$

where $\lambda^{(m)}$ will henceforth be used to denote the pertinent eigenvalue appropriate to the $m$th region, as defined in eqs (3.15) and (3.16). To find $Xe$ in eq (3.26), it is realized that eq (3.24) leads, via eqs (3.12) and (3.25), to

$$\left[A_{1-r}+\frac{B_{1-r}^2h^{4(r-1)}}{X}\right]\left[\overline{A}_{1-r}+\frac{\overline{B}_{1-r}^2h^{4(r-1)}}{\overline{X}}\right]h^4=1 \tag{3.27}$$

and the last relation is recognized to be a quadratic equation for $Xe$. By solving it and substituting into eq (3.26), one finally obtains:

$$\lambda_{\pm}^{(2r-1)}=(2r-1)^2+\Delta^2-G'e\cdot h^4\ \pm\sqrt{[\tilde{F}'\cdot h^{2(2r-1)}]^2+[2(2r-1)\Delta-G'o\cdot h^4]^2} \tag{3.28}$$

where

$$F'=\frac{B_{1-r}^2}{1-\bar{A}_{1-r}^2h^4}$$

$$G'=K_1+C_{1-r}-\overline{A}_{1-r}F'h^{4(r-1)}. \tag{3.29}$$

Result (3.28) is the required iteration formula for $m=2r-1$; though more complicated, this result is analogous to the iteration in eq (3.4) for $m=0$. The $\pm$ signs in $\lambda_{\pm}^{(m)}$ refer to the two admissible eigenvalues for given $h^2$ and $\Delta$. As already discussed in conjunction with the stability chart, all $\lambda^{(m)}$ $(m\neq 0)$ are double-valued for a given $\nu$ and $h^2$ in the unstable regions; the + and − signs then refer to the larger and the smaller values of $\lambda^{(m)}$, respectively. In the stable bands, $\lambda^{(m)}$ is single-valued and the two eigenvalues in eq (3.28) arise due to the choice of $m$ in eq (3.15); it is then easy to verify that $\lambda_{\pm}^{(m)}$ correspond to $\beta\gtrless 0$, i.e., $\lambda_{+}^{(m)}$ applies if $m<\nu<m+\frac{1}{2}$, whereas $\lambda_{-}^{(m)}$ is pertinent if $m-\frac{1}{2}<\nu<m$.

(b) $m$ even: $m=2r$ (r=1, 2, 3 . . . .)

For even values of $m$, eq (2.4) is written, with $n=-r$ and relations (3.11), (3.18), and (3.22), as

$$L_{-r}=(K_{1-r}+\bar{K}_{1-r})h^4 \tag{3.30}$$

which, by employing the notation of eqs (3.12) and (3.26), becomes

$$L_{-r}-2Ae_{1-r}h^4=\left(\frac{B_{1-r}^2}{X}+\frac{\bar{B}_{1-r}^2}{\bar{X}}\right)h^{4r}. \tag{3.31}$$

One then obtains again a quadratic with respect to $Xe$ which yields a similar, but different, solution when compared with the $m$ odd case. By noting the value of $L_{-r}$, as defined via eq (2.3), one now obtains

$$\lambda_{\pm}^{(2r)}=4r^2+\Delta^2-G''eh^4\pm\sqrt{(\tilde{F}''h^{4r})^2+(4r\Delta-G''o\cdot h^4)^2} \tag{3.32}$$

where

$$F''=\frac{B_{1-r}^2}{\Delta^2-\lambda_{\pm}^{(2r)}-2Ae_{1-r}h^4} \tag{3.33}$$

$$G''=K_1+C_{1-r}+F''h^{4(r-1)}. \tag{3.34}$$

Equation (3.32) is the required iteration formula for even values of $m$; the significance of the $\pm$ signs in $\lambda_{\pm}^{(m)}$ was already discussed in the paragraph following eq (3.29).

## 4. Formulas for the Eigenvalue $\lambda^{(m)}$ for Small $h^2$

Relations (3.28) and (3.32) are directly applicable for obtaining $\lambda^{(m)}$ to any desired accuracy by means of numerical methods. In addition, analytical formulations may be obtained for $\lambda^{(m)}$ provided $h^2$ is sufficiently small, this being particularly significant in certain physical situations. Thus, small $h^2$ occurs in problems wherein the periodicity with respect to the independent variable $\theta$ appears as a perturbation in an otherwise uniform configuration ($h^2=0$); it is then clear (e.g., from the stability chart) that instability phenomena may appear in certain ranges even if the perturbation $h^2$ is exceedingly small.

To examine these effects, all of the functions appearing in eqs (3.28) and (3.32) are expanded as power series via relations which are analogous to eq (3.6) and (3.9). Using considerations similar to those described in the paragraph following eq (3.9), formulas (3.28) and (3.32) are then employed to yield relations for $\lambda^{(m)}$ up to any desired power of $h^2$. This method is applied in the present section to obtain expansion for $\lambda^{(m)}$ for small $h^2$ up to and including $O(h^4)$. It is easy to verify that, if all terms of $O(h^6)$ or smaller are neglected, it is sufficient to retain only first terms ($h^0$) in the expansions for $K_1$, $A_n$, etc.; the approximations involved are commented upon in section 8. The following results are thus obtained:

(a) In the stable regions ($\nu$ real).

Formulas (3.4), (3.28) and (3.32) are used and expanded to $O(h^4)$; the iterations need be employed only once and with a trial value of $\lambda_0^{(m)}=\nu^2$. The exact definition of $\bar{L}_n$ is important in these derivations, as commented upon in the two paragraphs following eq (3.22).

$$m=0: \qquad 0\leq\nu\leq\frac{1}{2}; \qquad \lambda^{(0)}=\nu^2-\frac{h^4}{2(1-\nu^2)}+O\left(\frac{7}{128}h^8\right), \tag{4.1}$$

$$m=1: \qquad \frac{1}{2}\leq\nu_-\leq 1,\ 1\leq\nu_+\leq\frac{3}{2};$$

$$\lambda_{\pm}^{(1)}=1+(1-\nu_{\pm})^2-\frac{(5-\nu_{\pm})h^4}{16(1+\nu_{\pm})(2-\nu_{\pm})}$$

$$\pm\sqrt{h^4+\left[2+\frac{3h^4}{16(1+\nu_{\pm})(2-\nu_{\pm})}\right]^2(1-\nu_{\pm})^2}\pm O\left(\frac{h^6}{64}\right), \tag{4.2}$$

$m=2$: $\frac{3}{2}\leqslant \nu_{-}\leqslant 2,\ 2\leqslant \nu_{+}\leqslant \frac{5}{2}$;

$$\lambda_{\pm}^{(2)}=4+(2-\nu_{\pm})^2+\frac{(7-\nu_{\pm}^2)h^4}{(1-\nu_{\pm}^2)(3-\nu_{\pm})}$$

$$\pm\sqrt{\frac{h^8}{16(1-\nu_{\pm})^2}+\left[4+\frac{h^4}{6(1+\nu_{\pm})(3-\nu_{\pm})}\right]^2(2-\nu_{\pm})^2}-O\left(\frac{379\pm 384}{13824}h^8\right), \tag{4.3}$$

$m\geqslant 3$: $m-\frac{1}{2}\leqslant \nu_{-}\leqslant m,\ m\leqslant \nu_{+}\leqslant m+\frac{1}{2}$

$$\lambda_{\pm}^{(m)}=\nu_{\pm}^2-\frac{h^4}{2(1-\nu_{\pm}^2)}+O\left[\frac{5m^2+7}{32(m^2-1)^3(m^2-4)}h^8\right]. \tag{4.4}$$

It is noted that, to $O(h^4)$, only the results for $m=1$ and $m=2$ are different from eq (1.3). On the other hand, eqs (4.2) and (4.3) imply that eq (1.3) may lead to very erroneous results even for small perturbations; this relates to the regions which are close to $\nu=1$ and $\nu=2$, respectively.

The order of magnitude of the first neglected term in the above expansions is also indicated; the figures give the exact result for $\nu=m$ and it is expected that they yield a very good approximation for $\nu\neq m$.

(b) In the unstable regions ($\nu=m+i\alpha$).

As illustrated in figure 1, $h^2$ and $\nu$ may not be prescribed in a completely arbitrary fashion since, for $m\neq 0$ and a given $h^2$, $\alpha$ cannot exceed a certain value if $\lambda^{(m)}$ is to be real. The limiting $\alpha$ corresponds to the minimum of the constant-$\nu$ curves in the unstable region; this value of $\alpha$ may be calculated by stipulating that the radicands in eqs (3.28) and (3.32) vanish, thus obtaining

$$\alpha\leqslant w_m h^{2m}+O(h^{2m+2}) \tag{4.5}$$

where the equal sign refers to the limiting value and

$$w_m=\frac{1}{2m[2^{m-1}(m-1)!]^2}\ \text{(for all } m\neq 0). \tag{4.6}$$

Since the coefficients $w_m$ form a descending sequence, $\alpha$ is restricted to exceedingly small values for large $m$. This behavior is directly related to the extent of instability since $\alpha$ refers to the exponential increase of the solution (1.2); consequently, large values of $m$ (and therefore of $\lambda^{(m)}$) can accommodate only "weak" instabilities (at least for sufficiently small values of $h^2$).

With restriction (4.5) being accounted for, $\lambda^{(m)}$ may be obtained in the same fashion as in the stable regions, except that the trial values for the iteration are taken as $\lambda_0^{(0)}=-\alpha^2(m=0)$ and $\lambda_0^{(m)}=m^2(m\neq 0)$. It is then obtained:

$m=0$: $$\lambda^{(0)}=-\alpha^2-\frac{h^4}{2(1+\alpha^2)}+O\left[\frac{(7-5\alpha^2)h^8}{32(1+\alpha^2)^3(4+\alpha^2)}\right] \tag{4.7}$$

$m=1$: $$\lambda_{\pm}^{(1)}=1-\alpha^2-\frac{(8-\alpha^2)h^4}{(4+\alpha^2)(16+\alpha^2)}\pm\sqrt{h^4-(2\alpha)^2\left[1+\frac{3h^4}{(4+\alpha^2)(16+\alpha^2)}\right]^2}\mp O\left(\frac{h^6}{64}\right) \tag{4.8}$$

$m=2$: $$\lambda_{\pm}^{(2)}=4-\alpha^2+\frac{2(12+\alpha^2)h^4}{(4+\alpha^2)(36+\alpha^2)}\pm\sqrt{\left(\frac{h^4}{4+\alpha^2}\right)^2-(4\alpha)^2\left[1+\frac{2h^4}{(4+\alpha^2)(36+\alpha^2)}\right]^2}$$
$$-O\left(\frac{379\pm 384}{13824}h^8\right) \tag{4.9}$$

$m\geqslant 3$: $$\lambda_{\pm}^{(m)}=m^2+\frac{h^4}{2(m^2-1)}+O\left[\frac{5m^2+7}{32(m^2-1)^3(m^2-4)}h^8\right]. \tag{4.10}$$

Due to analogous considerations, the comments following eq (4.4) (with respect to $m \geq 3$) also apply here.

## 5. Characteristic Curves for the Periodic Mathieu Functions ($\nu = m$)

As a particular case, formulas (3.28) and (3.32) are applicable for obtaining the characteristic curves of the *periodic* Mathieu functions ($\nu = m$). Observing that $\Delta \equiv 0$ in these cases, all of the odd functions $Ao_n$, $Bo_n$, etc., are identically zero and the formulas simplify considerably.

The notation for the band edges is generally given as $a_m = \lambda_+^{(m)}(\nu) \mid_{\nu=m}$ and $b_m = \lambda_-^{(m)}(\nu) \mid_{\nu=m}$ in the literature [3–6]. One then obtains

$$a_0 = -2K_1 h^4, \tag{5.1}$$

$$b_{2r-1} = (2r-1)^2 - (K_1 + C_{1-r})h^4 - \frac{B_{1-r}^2 h^{2(2r-1)}}{1 + A_{1-r}h^2}, \tag{5.2}$$

$$a_{2r-1} = (2r-1)^2 - (K_1 + C_{1-r})h^4 + \frac{B_{1-r}^2 h^{2(2r-1)}}{1 - A_{1-r}h^2}, \tag{5.3}$$

$$b_{2r} = 4r^2 - (K_1 + C_{1-r})h^4, \tag{5.4}$$

$$a_{2r} = 4r^2 - (K_1 + C_{1-r})h^4 + \frac{B_{1-r}^2 h^{4r}}{\frac{1}{2}a_{2r} + A_{1-r}h^4}, \tag{5.5}$$

$$(r = 1, 2, 3 \ldots).$$

These relations may be used to generate the Mathieu function expansions for the eigenvalues $a_m$ or $b_m$ in terms of a power series in $h^2$. It is then evident that certain known features are already apparent in formulas (5.1)–(5.5); thus, odd values of $m$ lead to expansions in terms of powers of $h^2$ whereas even values of $m$ lead to expansions in terms of powers of $h^4$. In fact, if these expansions are obtained iteratively with a first trial value of $\lambda_0 = m^2$, the above formulas yield results which are identical with expressions available in the literature [3–7]. The identification follows upon recognizing that, for example, eqs (5.2)–(5.5) are a restatement of eqs (20–2, 21–24), respectively, on p. 723 of reference 6.

The advantage of the present formulation is evident in that it represents a systematic technique for generating the power expansions for $\lambda^{(m)}$. Whereas it is recognized that all methods (including the present one) become quite involved and tedious if $a_m$ or $b_m$ need to be found in terms of large powers of $h^2$, it is nevertheless clear that formulas (5.1)–(5.5) are already expressed in such a concise and systematic manner that considerable simplification occurs.

## 6. Instability Intervals

The width of the band between each pair of the characteristic $\nu = m$ curves is of special importance since it yields information on the extent of the range of instability. For an arbitrary value of $h^2$, this instability interval is defined by

$$W_m = a_m - b_m \quad (m \neq 0) \tag{6.1}$$

and is easily determined via relations (5.1) to (5.5) which yield:

$$W_{2r-1} = \frac{2B_{1-r}^2}{1 - A_{1-r}^2 h^4} h^{2(2r-1)} \tag{6.2}$$

$$W_{2r} = \frac{2B_{1-r}^2}{a_{2r} + 2A_{1-r}h^4} h^{4r} \tag{6.3}$$

Both results may be expressed also as

$$W_m = 2\left[\frac{h^m}{2^{m-1}(m-1)!}\right]^2 + O(h^{2m+2}) \tag{6.4}$$

which is obtained by taking the first term in a series expansion of eqs (6.2) or (6.3) in powers of $h^2$. It is observed that, to this order of approximation

$$W_m \cong 4mw_m h^{2m} \tag{6.5}$$

where $w_m$ are the coefficients in eq (4.6). Result (6.4) is identical to that already obtained by Bell [17], Levy and Keller [18] and Hochstadt [19] who used different approaches. In the present case, the instability interval $W_m$ turns out as a simple corollary to formulas (5.2)–(5.5) and the accuracy of eq (6.4) can be estimated as discussed in section 8. It is also noted that $W_m$ is simply related to the maximum allowable value of $\alpha$ since, introducing eq (4.5) into eq (6.5), one has

$$\alpha \leq \frac{W_m}{4m} + O(h^{2m+2}). \tag{6.6}$$

Hence, as $m$ increases, the range of $\alpha$ becomes more restricted than that of $W_m$. Phrased differently, larger values of $\nu$ (or $\lambda$) lead to regions with decreasing unstable intervals; the extent of instability itself is proportional to the interval $W_m$, but inversely proportional to $\nu$ (or $\lambda$).

## 7. Magnitude of Harmonics

The magnitude of the harmonic coefficients $c_n(\pm i\nu)$ of eq (1.2) may be found by using the appropriate values of $\lambda^{(m)}$ obtained above. For simplicity, the + sign in $c_n(\pm i\nu)$ will be implied subsequently; hence, $c_n$ is taken to denote $c_n(i\nu)$. To find the coefficients $c_n(-i\nu)$, one may employ the identity $c_n(-i\nu) = c_{-n}(i\nu)$ which is proven in the literature [3–6].

Using eqs (2.1), (2.2), (3.11), and (3.22), one gets

$$\frac{c_n}{c_{n-1}} = -h^2 K_n \tag{7.1}$$

$$\frac{c_n}{c_{n+1}} = \frac{\bar{c}_{-(m+n)}}{\bar{c}_{-(m+n+1)}} = -h^2 \bar{K}_{-(m+n)} \tag{7.2}$$

As clarified below, the $c_0$ and $c_{-m}$ harmonics turn out to be particularly important. For these harmonics, one obtains the following relations:

(a) For $m$ odd; $m = 2r-1 (r = 1, 2, 3 \ldots)$.

By using eq (7.1) successively, one obtains

$$\frac{c_0}{c_{1-r}} = \prod_{i=0}^{r-2} \frac{c_{-i}}{c_{-1-i}} = (-h^2)^{r-1} \prod_{i=0}^{r-2} K_{-i} \tag{7.3}$$

and eq (7.2) similarly yields:

$$\frac{c_{1-2r}}{c_{1-r}} = \prod_{i=r}^{2r-1} \frac{c_{-i}}{c_{1-i}} = (-h^2)^r \cdot \prod_{i=0}^{r-1} \bar{K}_{-i} \tag{7.4}$$

so that, by division:

$$R' = \frac{c_{1-2r}}{c_0} = \begin{cases} -h^2 \bar{K}_0 \text{ (for } r = 1) \\ \\ -h^2 \bar{K}_{1-r} \prod_{i=0}^{r-2} \dfrac{\bar{K}_{-i}}{K_{-i}} \text{ (for } r > 1) \end{cases} \tag{7.5}$$

(b) For $m$ even; $m=2r(r=1, 2, 3\ .\ .\ .)$.

In a manner similar to that for $m$ odd, one gets:

$$\frac{c_0}{c_{-r}}=\prod_{i=0}^{r-1}\frac{c_{-i}}{c_{-1-i}}=(-h^2)^r\prod_{i=0}^{r-1}K_{-i} \tag{7.6}$$

$$\frac{c_{-2r}}{c_{-r}}=\prod_{i=r+1}^{2r}\frac{c_{-i}}{c_{1-i}}=(-h^2)^r\prod_{i=0}^{r-1}\bar{K}_{-i} \tag{7.7}$$

and, by division:

$$R''=\frac{c_{-2r}}{c_0}=\prod_{i=0}^{r-1}\frac{\bar{K}_{-i}}{K_{-i}}. \tag{7.8}$$

By using eqs (7.1), (7.2), (7.5), and (7.8), one may now express all $c_n$ by means of the following relations:

$$\frac{c_n}{c_0}=\begin{cases}(-h^2)^n\prod_{i=1}^{n}K_i & (n>0)\\ (-h^2)^{-n}\prod_{i=m+n}^{m-1}\bar{K}_{-i} & (-r<n<0)\\ (-h^2)^{m+n}R\prod_{i=-n}^{m-1}K_{-i} & (-m<n\leqslant -r)\\ (-h^2)^{-(m+n)}R\prod_{i=1}^{-(m+n)}\bar{K}_i & (n<-m)\end{cases} \tag{7.9}$$

and it is understood that $R$ stands for $R'$ or $R''$ for odd or even values of $m$, respectively. The fundamental harmonic $c_0$ may be taken as unity or normalized via an orthogonal relationship [3−6]; all these normalization procedures, however, lead to a real value for $c_0$.

Equations (7.5) and (7.8) lead to the useful relation

$$\tilde{R}^2=R\bar{R}=1 \tag{7.10}$$

which is obvious for $R=R''$; for $R=R'$, this is verified by observing that eq (3.24) accounts for the extra $K_{1-r}$ term in eq (7.5). One therefore gets

$$\tilde{c}_n^2=\tilde{c}_{-(m+n)}^2 \tag{7.11}$$

which is obtained via eq (7.9) and (7.10).

Result (7.11) is interesting since one recalls that, in the unstable bands, any function $F(\nu)$ leads to a complementary function $F(\bar{\nu})=\bar{F}(\nu)$ which is identical to $F^*(\nu)=F(\nu^*)$, where the asterisk denotes the complex conjugate. Consequently, the modulus $\tilde{F}(\nu)$ is then identical to the absolute value $|F(\nu)|$, so that

$$|c_n|=|c_{-(m+n)}| \tag{7.12}$$

in the unstable bands. The harmonic amplitudes are then equal in pairs, as shown in figure 2. Relation (7.12) holds up to and including the band edges ($\nu=m$) where all variables are real; hence

$$c_n=\pm c_{-(m+n)} \tag{7.13}$$

since all terms are real and therefore any function $\bar{F}(\nu)=F^*(\nu)=F(\nu)$ at the band edges. The proper sign in eq (7.13) is determined by the sign of $R=\pm 1$ itself at the band edges. For $m$ odd, the sign of $R'$ is given by the sign of $-h^2\bar{K}_{1-r}=-h^2K_{1-r}$ because the product term in eq (7.5) is obviously $+1$ at the band edge. By using eqs (3.12), (3.13), and (3.24), one has

$$-h^2K_{1-r}=-A_{1-r}h^2-\frac{B_{1-r}^2h^{2(2r-1)}}{(2r-1)^2-\lambda_{\pm}^{(2r-1)}-(K_1+C_{1-r})h^4}=R'=\pm 1 \tag{7.14}$$

where $\lambda_{-}^{(2r-1)}$ refers to $b_{2r-1}$ and $\lambda_{+}^{(2r-1)}$ refers to $a_{2r-1}$ of eqs (5.2) and (5.3), respectively. It is easily seen by inspection that, in order to comply with eqs (5.2) and (5.3), $R'=+1$ need be taken for $a_{2r-1}$ whereas $R'=-1$ is appropriate for $b_{2r-1}$ in eq (7.14).

The even $m$ case is somewhat more complex. Checking for $K_{1-r}$, it is seen that $a_{2r}$ of eq (5.5) leads to a finite value for $K_{1-r}$ but $b_{2r}$ of eq (5.4) requires that $|K_{1-r}|\to\infty$. Hence $R''=+1$ for $a_{2r}$ (since then all $K_{-i}=\bar{K}_{-i}$ are finite) but is yet undetermined for $b_{2r}$. To clarify the latter situation, one uses eq (3.30) to find the limit

$$\lim_{\Delta\to 0}\frac{\bar{K}_{1-r}}{K_{1-r}}=\lim_{\Delta\to 0}\left[\frac{L_{-r}}{h^4K_{1-r}}-1\right]=-1$$

and since all $K_{-i}=\bar{K}_{-i}$ $(i\neq r-1)$ in eq (7.8) at the edge pertinent to $b_{2r}$, eq (7.15) leads to $R''=-1$ at that edge. In conclusion, one obtains that

$$\begin{aligned} c_n&=+c_{-(m+n)} \text{ for } a_m \\ c_n&=-c_{-(m+n)} \text{ for } b_m \end{aligned} \tag{7.15}$$

and therefore the function $y(\theta)$ of eq (1.2) is in the form of a Fourier cosine or sine expansion, respectively, on the right or left edge of the unstable regions.

A special feature of the $b_{2r}$ edge is that, since $K_{1-r}\to\infty$, one gets from eq (7.6) that $c_{-r}=0$ which also agrees with the second of eqs (7.15); hence, the $c_{-r}$ harmonic vanishes at the left edge of all the unstable regions which are characterized by even values of $m=2r$. It is appreciated that this harmonic is exceptional in that it is the only one which (for $m$ even) does not pair off with any other harmonic in the unstable regions but, as shown in figure 2 and eq (7.12), it can be said to pair with itself.

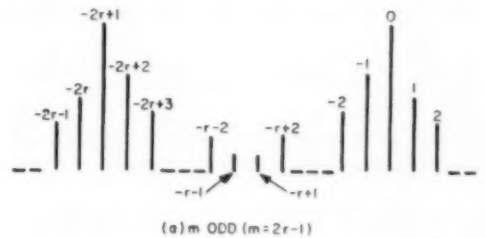
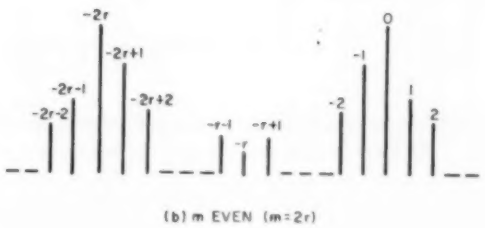



FIGURE 2. *The magnitudes of the harmonic coefficients* $c_n(iv)$ *in the unstable regions.*



To conclude the discussion, it is pertinent to find the magnitudes of the harmonics for small $h^2$. In that case, one may approximate $K_i$ in the unstable regions as

$$\frac{1}{K_i} \cong L_i \cong (m+\Delta+2i)^2-(m+\Delta)^2 \cong 4i(m+i) \cong \frac{1}{\overline{K}_i}. \tag{7.16}$$

For $K_{-i}$, however, condition (4.5) precludes an easy approximate evaluation. The order of magnitude may nevertheless be estimated by replacing $i$ with $-i$ in eq (7.16). One then finds:

$$\left|\frac{c_n}{c_0}\right| = \begin{cases} \left(\frac{h^2}{4}\right)^n \frac{m!}{n!(m+n)!} + O(h^{2n+2}) & (n>0) \\ \left(\frac{h^2}{4}\right)^{-n} O\left[\frac{(m+n-1)!}{(-n)!(m-1)!}\right] & (-r<n<0) \\ \left(\frac{h^2}{4}\right)^{m+n} O\left[\frac{(-n-1)!}{(m+n)!(m-1)!}\right] & (-m<n\leq -r) \\ \left(\frac{h^2}{4}\right)^{-(m+n)} \cdot \frac{m!}{(-m-n)!(-n)!} + O(h^{2-m-n}) & (n\leq -m) \end{cases} \tag{7.17}$$

for the unstable regions. Numerical and approximate analytical considerations indicate that, for $m>3$, the order of approximation for $-m<n<0$ in eqs (7.17) approaches the actual value very closely.

In the stable regions, property (7.10) does not hold any longer and $\overline{K}_{-i}$ turns out to be quantitatively dissimilar from $K_{-i}$. To find the latter, one notes that eqs (3.1) and (3.13) imply that

$$S-C_{-i}h^4 = L_0-K_1h^4-C_{-i}h^4 = (K^- - C_{-i})h^4 \tag{7.18}$$

where $K^-$ and $K^+=K_1$ were defined in eq (3.3). Employing the definition of $C_{-i}$ in eq (3.14) and the series expansion (3.6) for $K^-$, one finds

$$K^- - C_{-i} = B_{-i}B_{-(i+1)}h^{4i} + O[h^{4(i+1)}]. \tag{7.19}$$

Introducing eqs (7.18) and (7.19) into eq (3.12), one obtains via eqs (3.7) and (3.8)

$$K_{-i} = A_{-i} + \frac{1}{A_{-(i+1)}h^4} + O(h^0) = \frac{L_{-(i+1)}}{h^4} + O(h^0) = \frac{4(i+1)(i+1-\nu)}{h^4} + O(h^{-2}). \tag{7.20}$$

In view of the comments following eq (3.22), $\overline{K}_{-i}$ cannot be obtained directly from eq (7.20) and one needs to use definition (3.12) instead. It is then observed that

$$\overline{S} = \overline{L}_0 - \overline{K}_1h^4 = (2m-\nu)^2 - \lambda - \overline{K}_1h^4; \tag{7.21}$$

also, an inspection of eqs (4.1)–(4.4) for $\lambda$ reveals that $\overline{S}$ is of $O(h^0)$, provided $\nu$ is not too close to $m$ (band-edge). Thus

$$\overline{K}_{-i} = \overline{A}_{-i} + O(h^{4i}) = \frac{1}{\overline{L}_{-i}} + O(h^4) = \frac{1}{4(m-i)(m-\nu-i)} + O(h^2) \qquad (i \neq m) \tag{7.22}$$

and this approximation is valid for use in eqs (7.9) since $i \neq m$ therein.

The terms $K_i$ and $\overline{K}_i$ are obtained in a similar form to that of the unstable regions. Thus,

$$K_i = \frac{1}{L_i} + O(h^4) = \frac{1}{4i(i+\nu)} + O(h^2), \tag{7.23}$$

$$\bar{K}_i = \frac{1}{L_i} + O(h^4) = \frac{1}{4(m+i)(m-\nu+i)} + O(h^2). \tag{7.24}$$

Introducing the relations above into eqs (7.9) one has for the stable regions:

$$\frac{c_n}{c_0} = \left(-\frac{h^2}{4}\right)^{|n|} \frac{\Gamma(1 \pm \nu)}{|n|!\Gamma(1 + |n| \pm \nu)} + O[h^{2(1+|n|)}] \quad (\text{for } n \gtrless 0) \tag{7.25}$$

where $\Gamma(x)$ is the gamma function of argument $x$.

When comparing the last results with eqs (7.17), one recognizes that the order in $h^2$ for $|c_n/c_0|$ is the same in both stable and unstable regions for $n \geqslant -r$; for $n < -r$, on the other hand, $|c_n/c_0|$ changes from $O(h^{2|m+n|})$ to $O(h^{2|n|})$, i.e., by a factor of $h^{2m}$. This is due to the fact that pairing of the harmonics occurs in the unstable region only.

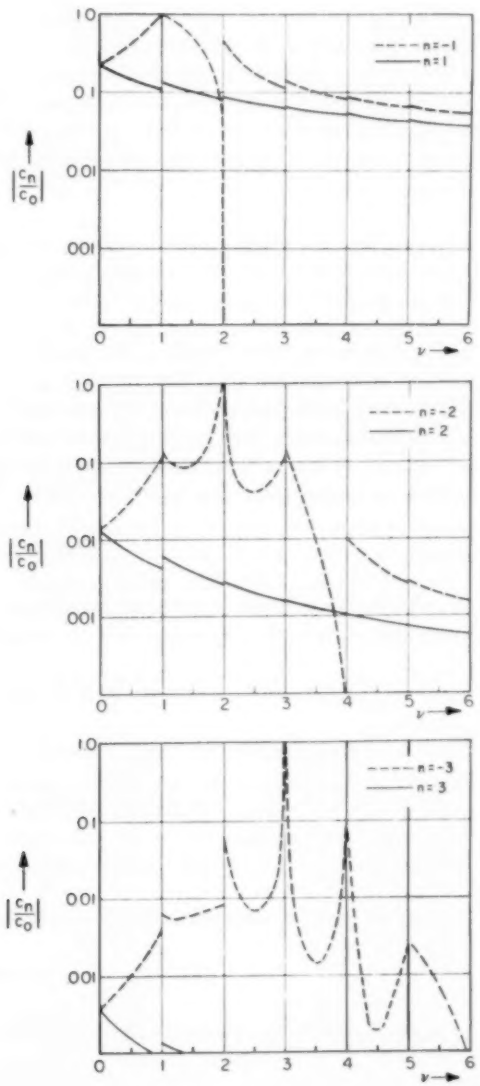



FIGURE 3. *Variation of the harmonic coefficients* $c_n(i\nu)$ *within the stable regions, for* $h^2 = 1$.

The above considerations indicate the following: (a) for $n \geq -r$, $|c_n|$ decreases monotonically relative to $c_0$ as $\lambda$ increases; (b) every harmonic $|c_n|$ with $n < -r$ increases near and at the band edges and decreases towards the middle of the stable bands. These remarks apply to cases with $h^2$ small and, as one follows the value of $c_n$ with $n < 0$, it is clear that this harmonic will behave in the manner (b) up to a value of $\lambda \cong r$ and in the manner (a) for $\lambda > r$. For $n > 0$, the harmonic amplitudes behave only in the manner (a) since then $n$ is always greater than $-r$.

The above features are illustrated in figure 3 wherein a few harmonic amplitudes are plotted for the range $0 \leq \nu \leq 6$ within the stable regions. The vanishing of the $c_{-r}$ harmonic at the left band edge with $m$ even is shown in the figure by $c_{-1}$, $c_{-2}$, and $c_{-3}$ which vanish respectively, at $\nu = 2$, 4 and 6. Also, relations (7.16) are verified; thus, $|c_{-1}| = c_0$ at $\nu = 1$, $|c_{-2}| = c_0$ at $\nu = 2$, $|c_1| = |c_{-2}|$ at $\nu = 3$, etc. The constant monotonic behavior of $c_n$ for $n > 0$ should also be observed.

## 8. A Note on Accuracy and Convergence

In the foregoing derivations, the exact meaning of "$h^2$ small" was not explored and this question is examined in the present section.

It is first remarked that, if a numerical computation is desired, the results (3.28) and (3.32) may be valid for the calculation of $\lambda$ for any values of $\nu$ and $h^2$. One notes that all the expressions involved are in the form of *finite* continued fractions, except for $K_1 = K^+$ which is infinite. The latter, however, is expandable via eq (3.6) into a series consisting of terms $\kappa_k$ that, for large $k$, vary as

$$\kappa_k = B_k B_{k+1} h^{4k} \rightarrow \frac{h^{4k}}{L_{k+1} \prod_{j=1}^{k} L_j} \rightarrow \frac{1}{4(k+1)(m+k+1)} \left[\frac{m!}{k!(m+k)!}\right]^2 \left(\frac{h}{2}\right)^{4k} \tag{8.1}$$

which converges very rapidly. Hence, for numerical computation, $\lambda$ is obtainable within any desired accuracy provided the iteration process itself converges. However, the convergence of the iteration process is still an open question.

With the above observations in mind, the accuracy of the results for small $h^2$ can be examined. As already noted, only the leading terms in the expansions for $K_1$, $A_{1-r}$, etc., need be retained for results up to and including a term of $O(h^4)$. With respect to $K_1$, this means that one needs

$$h^4 \ll |L_i L_{i+1}| \text{ for all } i > 0, \tag{8.2}$$

in order to justify the assumed approximation. All of the other functions $A_{1-r}$, $B_{1-r}$, and $C_{1-r}$ are essentially expressible in terms of $A_{1-r}$; retaining only the first term in the latter means that

$$h^4 \ll |L_{-j} L_{1-j}| \text{ for all } 1 < j < r \tag{8.3}$$

If one replaces $\lambda$ by the first approximation $\lambda_0 = \nu^2 \cong m^2$ in the two inequalities, one gets:

$$h^4 \ll 16i(i+1)(m+i)(m+i+1) \qquad (i > 0) \tag{8.4}$$

and

$$h^4 \ll 16j(j-1)(m-j)(m-j+1) \qquad (1 < j < r). \tag{8.5}$$

The worst case for inequality (8.4) is $m = 0$, $i = 0$, i.e.,

$$h^4 \ll 64 \tag{8.6}$$

whereas, for (8.5), the worst restriction occurs at $m = 5$ and $j = 2$ since only $1 < j < r$ applies; hence,

$$h^4 \ll 12 \times 32. \tag{8.7}$$

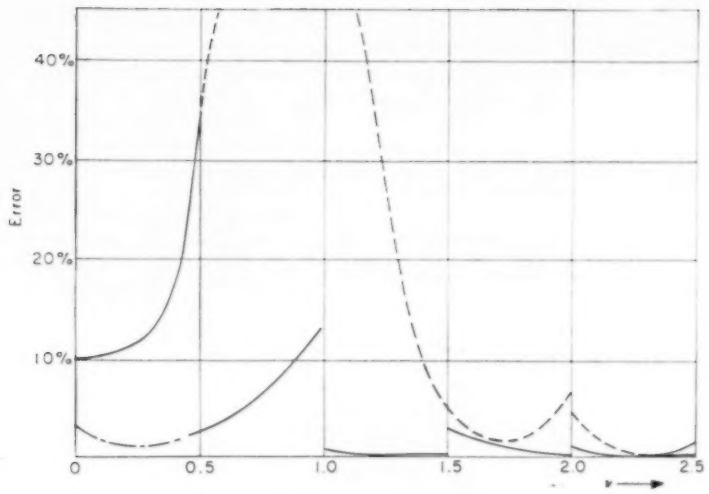



FIGURE 4. *Percentage error incurred in the value of* $\lambda$, *for* $h^2 = 1$.
Solid lines refer to the error produced by eqs (4.1)–(4.4) and dashed lines refer to the error due to the use of eq (1.3). The dot-dashed curve relates to the error which occurs when eq (4.2) is used instead of eq (4.1).

Equations (8.6) and (8.7) indicate that the results obtained should hold well up to and including values of $h^2$ of order unity. An analogous examination of $\overline{L}$ (instead of $L$) leads to the same restriction. To check these considerations, the exact results for $h^2 = 1$ are compared in figure 4 with those obtained via eqs (4.1)–(4.4). The solid lines indicate the error produced by using eqs (4.1)–(4.4) whereas the dashed lines refer to the results obtained via eq (1.3). It is noted that considerable improvement is obtained, especially near $\nu = 1$ and $\nu = 2$. It is also interesting to observe that smaller errors are obtained if eq (4.2) is used to calculate $\lambda$ for $0 \leq \nu \leq 0.5$ instead of eq (4.1); this feature is shown in figure 4 by the dot-dashed line. Calculations were also carried out for $h^2 = 0.1$ and 2.5; the errors incurred then are respectively, less than 0.1 percent and 20 percent for all $\nu$, provided eqs (4.2)–(4.4) are employed (i.e., eq (4.2) is extended to cover the range $0 < \nu < 3/2$).

The above comments indicate that the formulation discussed in the present paper is useful from both analytical and computational aspects. While it is recognized that the iteration process may not always be convergent, it is also probable that the present formulation might provide a tool for examining the range of validity for some of the expressions of the Mathieu functions and solutions whose convergence is still an open question.

---

The authors are grateful to Professor A. A. Oliner for his encouragement and Professors E. S. Cassedy and H. Hochstadt for helpful suggestions; also, thanks are due to Dr. Gertrude Blanch, Wright-Patterson Air Force Base, Ohio for her criticism of the manuscript. This work was sponsored by the Air Force Cambridge Research Laboratory, Office of Aerospace Research, under contract No. AF–19 (604)–7499.

## 9. Appendix

Assume that $K_n$, as defined in eq (3.11), may be written as

$$K_n = P_n + \frac{Q_n}{S - R_n} \qquad (n \leq 0) \tag{A1}$$

where $P_n$, $Q_n$, and $R_n$ are functions to be determined in such a manner that the above becomes an identity; $S$ was defined in eq (3.13). One then has

$$K_{n-1}=\frac{1}{L_{n-1}-h^4K_n}\equiv P_{n-1}+\frac{Q_{n-1}}{S-R_{n-1}}. \tag{A2}$$

Substituting for $K_n$, eq (A2) yields after some algebraic manipulation,

$$P_{n-1}+\frac{Q_{n-1}}{S-R_{n-1}}\equiv\frac{1}{L_{n-1}-h^4P_n}\left[1+\frac{\dfrac{Q_n}{L_{n-1}-h^4P_n}}{S-R_n-\dfrac{Q_n}{L_{n-1}-h^4P_n}}\right] \tag{A3}$$

and, identifying terms, one obtains

$$P_{n-1}=\frac{1}{L_{n-1}-h^4P_n} \tag{A4}$$

$$Q_{n-1}=\frac{Q_nh^4}{(L_{n-1}-h^4P_n)^2} \tag{A5}$$

$$R_{n-1}=R_n+\frac{Q_nh^4}{L_n-h^4P_n}. \tag{A6}$$

The last relations form a system of difference equations which are solved by noting that the boundary condition at $n=0$ stipulates that

$$P_0=R_0=0 \text{ and } Q_0=1. \tag{A7}$$

Applying this to (A4)–(A6), one obtains easily

$$P_n=\frac{1}{L_n}\Big|-\frac{h^4}{L_{n+1}}\Big|-\frac{h^4}{L_{n+2}}\Big|-\cdots-\frac{h^4}{L_{-2}}\Big|-\frac{h^4}{L_{-1}} \tag{A8}$$

$$Q_n=h^{-4n}\prod_{i=1}^{-n}P_{-i}^2 \tag{A9}$$

$$R_n=h^4\sum_{j=1}^{-n}P_{-j}Q_{1-j} \tag{A10}$$

all of which hold for $n<0$ and it is understood that $P_{-1}=1/L_{-1}$. Upon identifying

$$A_n\equiv P_n, \qquad B_n^2\equiv Q_nh^{4n} \text{ and } C_n=R_nh^{-4}, \tag{A11}$$

it is readily verified that eq (3.12) is proven and that eqs (3.7), (3.8), and (3.14) are satisfied.

## 10. References


[1] E. Mathieu, J. de Math. Pures & Appl. (Liouville) **13,** 137 (1868).

[2] G. W. Hill, Acta Math. **8,** 1 (1886).

[3] N. W. McLachlan, Theory and Application of Mathieu Functions (Clarendon Press, Oxford, 1947).

[4] J. Meixner and F. W. Schäfke, Mathieusche Funktionen und Sphäroidfunktionen (Springer-Verlag, Berlin, 1954).

[5] R. Campbell, Theorie Generale de l'Equation de Mathieu (Masson and Co., Paris, 1955).

[6] National Bureau of Standards, Handbook of Mathematical Functions (U.S. Dept. of Commerce, 1964), ch. 20, p. 721, by Gertrude Blanch.

[7] National Bureau of Standards, Tables Relating to Mathieu Functions (Columbia University Press, N.Y., 1951).

[8] Lord Rayleigh, Phil. Mag. **24,** 145 (1887).
[9] L. Brillouin, Wave Propagation in Periodic Structures, ch. 8, p. 172 (Dover Publications, Inc., N.Y., 1953).
[10] T. Tamir, H. C. Wang and A. A. Oliner, IEEE Trans. on Microwave Theory and Tech. **MTT-12,** 323 (1964).
[11] N. Minorsky, Nonlinear Oscillators, ch. 16.4 and 20. (Van Nostrand Co., Inc., Princeton, N.J., 1962).
[12] T. Tamir, Math. Comput. **16,** 100 (1962).
[13] J. Meixner and F. Schäfke, op. cit., ch. 2.25, p. 119.
[14] F. W. Schäfke, R. Ebert and H. Groh, Num. Math., **4,** 1 (1962).
[15] B. van der Pol and M. J. O. Strutt, Phil. Mag. **5,** 18 (1928).
[16] O. Perron, Die Lehre von Kettenbrüchen, ch. II–45, p. 205. (Chelsea Publishing Co., N.Y., 1950).
[17] M. Bell, Proc. Glasgow Math. Assoc. **3,** 132 (1957).
[18] D. M. Levy and J. B. Keller, Comm. Pure and Appl. Math. **16,** 469 (1963).
[19] H. Hochstadt, Comm. Pure and Appl. Math. **17,** 251 (1964).

(Paper 69B1–141)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# On the Surface Duality of Linear Graphs*

**Jack Edmonds**

(February 3, 1964)

THEOREM: A 1-1 correspondence between the edges of two connected graphs is a duality with respect to some polyhedral surface embedding if and only if for each vertex $v$ of each graph, the edges which meet $v$ correspond in the other graph to the edges of a subgraph $G_v$ which is connected and which has an even number of its edge-ends to each of its vertices (where if an edge meets $v$ at both ends its image in $G_v$ is counted twice). Using the Euler formula, the characteristic of the surface is determined by the two graphs. Thus, the theorem generalizes a variation of the H. Whitney condition for a graph to be planar.

A graph, for purposes here, is a finite set of edges, i.e., homeomorphs of a closed line segment, and an identification of the edge endpoints into a number of equivalence classes called vertices. We call edge endpoints simply *ends* of the edge. We say that an edge end *meets* the vertex which is its image.

The edges and vertices of a polygonal disk, bounded by one or more edges, form a graph—a particular kind, called a circuit. When we identify in pairs, according to homeomorphisms, the edges of a finite set of polygonal disks, we form (by composing these identifications with those in the disks) one or more closed surfaces. The pair $S_U=(S, G_U)$, consisting of one such surface, $S$, and the graph, $G_U$, whose edges and vertices are the distinct images in $S$ of disk edges and vertices, is called a *polyhedral surface*, or a *map*, or a *polyhedral surface embedding of* $G$, where $G$ is a homeomorph of $G_U$.

The vertices and edges of the graph $G_U$ are also called the vertices and edges of $S_U$. The disks are called the faces of $S_U$. The graph $G_U$ of a map is connected in the usual sense. Each of its edges is the image of two face-edges. A vertex of $G_U$ is the image of one or more face vertices and also the image of the same number of edge ends of edges of $G_U$.

Associated with any map $S_U=(S, G_U)$ is a (topologically unique) dual map $S_W=(S, G_W)$ with the properties that (a) a map is the dual map of its dual map, (b) each vertex of a map is interior to a face of its dual map, (c) each face of a map contains exactly one vertex of its dual map, and (d) each edge of a map crosses once exactly one edge of the dual map.

By an *edge-correspondence* between two graphs, we mean a 1–1 correspondence between the edges of the two graphs. An edge-correspondence is called a *surface duality correspondence* or briefly a *duality* if the two graphs respectively have homeomorphisms with the graphs of dual maps so that the images of corresponding edges cross each other.

r main objective, Theorem 1, will give a necessary and sufficient condition for an edge-correspondence between two graphs $G_U$ and $G_W$ to be a surface duality correspondence.

Using Euler's formula, the characteristic $x$ of the surface for a surface duality correspondence is determined by the two graphs. Where $V_U$, $E_U$, and $F_U$ are the numbers of vertices, edges, and faces in map $S_U$ and $V_W$, $E_W$, and $F_W$ are the numbers of vertices, edges, and faces in dual map $S_W$, we have $x(S)=V_U-E_U+F_U=V_W-E_W+F_W$ and $V_U=F_W$, $E_U=E_W$, and $F_U=V_W$.

Hence,

$$x(S)=V_U-E_U+V_W. \tag{1}$$

COROLLARY 1 (to Theorem 1): A necessary and sufficient condition for a connected graph $G_U$ to have a polyhedral surface embedding of given characteristic $x$ is that it have an edge-correspondence with some graph $G_W$ such that (a) the condition of Theorem 1 is satisfied and (b) formula 1 is satisfied.

Corollary 1 is analogous to Hassler Whitney's well-known theorem on the planar duality of graphs[1]. He defines two graphs to be dual if there is an edge-correspondence between them that satisfies certain combinatorial conditions. His theorem states that a graph is planar if and only if it has a dual.

That a connected graph be planar is equivalent to its having a polyhedral surface embedding which is spherical. A polyhedral surface is a sphere if and only if it has characteristic equal two. Hence, for $x=2$ conditions (a) and (b) of Corollary 1 are equivalent (for connected graphs) to Whitney's notion of duality.

THEOREM 1: A 1-1 correspondence between the edges of two connected graphs is a duality with respect to some polyhedral surface embedding if and only if for each vertex $v$ of each graph, the edges which meet $v$ correspond in the other graph to the edges of

* Sponsored by the Office of Naval Research and the Army Research Office, Durham. Prepared while the author was a research associate at Princeton University.

[1] H. Whitney, Non-separable and planar graphs, Trans. Am. Math. Soc. **34**, 339–362 (1932).

a subgraph $G_v$ which is connected and which has an even number of its edge-ends to each of its vertices (where if an edge meets $v$ at both ends its image in $G_v$ is counted twice).

*Proof of necessity.* A graph $B_v$, formed by the edges and vertices of a polygonal disk $D_v$ before identifications in a map, is connected and has exactly two edge-ends meeting each of its vertices. In other words, $B_v$ is a *circuit* or a *simple closed path*. Under the identification which occurs when $D_v$ is regarded as a face of a map $S_U=(S, G_U)$, the image of $B_v$, call it $\hat{G}_v$, is a connected subgraph of $G_U$.

Each vertex $u$ in $G_v$ is the image of a set $V_{v,u}$ of one or more vertices of $B_v$. Each edge in $G_v$ is the image of either one or two edges in $B_v$; in the later case call it a *double edge* in $G_v$. Conversely, the image of every edge-end and vertex which meet in $B_v$ is an edge-end and vertex which meet in $G_v$. Hence, if we count twice the edge-ends of each double edge in $G_v$, then $u$ meets an even number of edge-ends of edges in $G_v$, two for each member of $V_{v,u}$.

Each edge of $G_v$ crosses exactly one edge of $G_W$ where $S_W=(S, G_W)$ is the dual of $S_U$. These edges of $G_W$ are exactly those which meet the one vertex $v$ of $G_W$ which is interior to disk $D_v$. Furthermore, one of these edges meets $v$ at both ends if and only if the edge of $G_v$ it intersects is double. Thus, the "only if" part of the theorem is proved.

A *closed path* in a graph $G$ is a mapping $C$ of a circuit $B$ into $G$ which takes edge-interiors of $B$ homeomorphically to edge-interiors of $G$ and which takes edges to edges and vertices to vertices so that edge-vertex incidence relations are preserved. Equivalently, a *closed path* is a cyclic order $(\ldots v_iE_iv_{i+1}E_{i+1}\ldots)$, $i$ taken modulo some $n$, of vertices and edges in $G$ such that for all $i$ edge $E_i$ meets vertex $v_i$ at one end and vertex $v_{i+1}$ at the other end. There is no distinguished position or orientation of the cyclic order. $E_i$ and $E_{i+1}$ are said to be *adjacent at* vertex $v_{i+1}$ in the cyclic order.

An *Euler path* in a graph $G$ is defined as a closed path which contains each edge of $G$ exactly once. (Here, a "double-edge" is to be regarded as two edges.)

Euler's "Konigsberg Bridge" theorem states that a graph $G$ has an Euler path if and only if $G$ is connected and has an even number of edge-ends to each vertex. Theorem 1 is really a strengthened form of Euler's theorem. It states conditions for the existence of a whole family of Euler paths which fit together in a certain way.

To construct in $G$, connected and even at each vertex, an Euler path, first construct a family of closed paths which together contain each edge of $G$ exactly once. Try doing this and you can't miss. Because $G$ is connected, if the family of paths has more than one member then some two members $C_a$ and $C_b$ have at least one vertex $v_0$ in common. If $C_a$ and $C_b$ can always be replaced in the family by one closed path, then by iteration an Euler path can be obtained.

Suppose $(\ldots E_1^av_0E_2^a\ldots)$ in $C_a$ and $(\ldots E_1^bv_0E_2^b\ldots)$ in $C_b$. There are *two ways* to change edge-adjacencies at $v_0$, either to $E_1^av_0E_1^b$ and $E_2^av_0E_2^b$ or to $E_1^av_0E_2^b$ and $E_2^av_0E_1^b$, so that with the other adjacencies unchanged $C_a$ and $C_b$ join to a single closed path. That is, of the three possible adjacencies for $E_1^a$ at $v_0$, with $E_2^a$ or $E_1^b$ or $E_2^b$, one yields two closed paths and two yield one closed path. This extra freedom in constructing an Euler path is important to Theorem 1.

*Proof of sufficiency.* Let there be an edge-correspondence between connected graphs $G_U$ and $G_W$ which satisfies the condition in Theorem 1. For each vertex $v$ in $G_W$, find an Euler path $C_v$ in $G_v$, where $G_v$ is the subgraph of $G_U$ which corresponds to the edges meeting $v$. Let each of these Euler paths bound a disk. More precisely, for each $v$ let the circuit, which is the pre-image of mapping $C_v$, bound a disk $D_v$; and then let the boundary of $D_v$ be identified with $G_v$ according to the mapping $C_v$. Because edges of the disks $D_v$ identify together in pairs according to continuous mappings that are 1–1 in their interiors, the resulting connected complex $K$ is locally planar except possibly at the vertices of $G_U$.

But what happens at a vertex $u$ of $G_U$? Each vertex of each disk $D_v$ meets two edge-ends in the disk and these edge-ends (for all of the disks $D_v$) identify in pairs to edge-ends of $G_U$ in $K$. This causes disk-vertices (or disk-vertex "corners") and the edge-ends of $G_U$ to arrange themselves in a number of cyclic orders. Let $H_u$ be the set of these cyclic orders whose disk-vertices are pre-images of $u$—that is, whose edge-ends meet $u$.

By construction of the cyclic orders, a pair of edges in $G_U$, with edge-ends adjacent at a disk vertex $d$ in cyclic order $h\epsilon H_u$, correspond to a pair of edges of $G_W$ which meet the vertex, say $v$, corresponding to the disk $D_v$ which contains $d$. Thus the edges of $G_U$ with edge-ends in $h\epsilon H_u$ correspond to a set of edges in $G_W$ which form a closed path $C_h$ in $G_W$ with cyclic order determined by $h$.

Graph $G_W$ can be embedded in $K$ so that each vertex $v$ in $G_W$ is contained in the interior of corresponding disk $D_v$ and so that edges of $G_W$ cross corresponding edges of $G_U$. Then $K$ is a map $(S, G_U)$ with dual $(S, G_W)$, if and only if for each $u$ of $G_U$ the family $H_u$ contains one member. Where $H_u$ contains $k_u$ cyclic orders, a neighborhood of $u$ in $K$ consists of $k_u$ open disks which intersect only at $u$. When $k_u=1$ for all $u$, the face of $(S, G_W)$ containing $u$ is a polygonal disk bounded by the pre-image of the mapping $C_h$, where $h$ is the only member of $H_u$.

The edges of $G_U$ which meet $u$ correspond to a subgraph, say $G_u$, of $G_W$. Since by hypothesis $G_u$ is connected, when $k_u>1$ there exist two closed paths $C_g$ and $C_h$, $g$ and $h\epsilon H_u$, which in $G_u$ have a common vertex, say $v$. Hence, closed path $C_v$ passes through $u$ at least twice, once at a vertex $d_g$ of $D_v$ where $d_g$ is a member of cyclic order $g\epsilon H_u$ and once at a vertex $d_h$ of $D_v$ where $d_h$ is a member of cyclic order $h\epsilon H_u$.

Suppose $(\ldots E_1^auE_1^b\ldots E_2^buE_2^a\ldots)$ describes the closed path $C_v$ where ends of edges $E_1^a$ and $E_1^b$ of $G_U$ are adjacent at $d_g$ in $g$ and where ends of edges $E_2^b$ and $E_2^a$ of $G_U$ are adjacent at $d_h$ in $h$. From $C_v$, by

changing the adjacencies $E_1^a u E_1^b$ and $E_2^b u E_2^a$ to $E_1^a u E_2^b$ and $E_1^b u E_2^a$ and leaving unchanged the other adjacencies in $C_v$, we obtain another Euler path $C_v'$ of $G_v$. Replacing disk $D_v$ by new disk $D_v'$, we get a new complex $K'$. At the same time, cyclic orders $g$ and $h$ combine to a single cyclic order with the same edge-end adjacencies except that, here also, the $E_1^a$, $E_1^b$ and $E_2^b$, $E_2^a$ adjacencies are replaced by $E_1^a$, $E_2^b$ and $E_1^b$, $E_2^a$ adjacencies.

The new family $H_u'$ of cyclic orders associated with $u$ contains one less cyclic order than $H_u$. The families of cyclic orders associated with the other vertices of $G_V$ remain the same. Repeating the process just described enough times we eventually get only one cyclic order in each family. Therefore, Theorem 1 is proved.

A corollary of Theorem 1 is that every connected graph $G$ can be embedded in some closed surface $S$ so that $S-G$ is an open disk. (For example, a circuit has such an embedding in the projective plane.)

Another corollary is that every connected graph $G$ which has an even number of edge-ends to each vertex is the graph of a map formed by identifying the edges of one polygonal disk with the edges of another. Notice how this corollary strengthens Euler's theorem—from asserting the existence of one Euler path in $G$ to asserting the existence of two Euler paths in $G$ (distinct unless $G$ is a circuit) which fit together in a certain way.

The characteristic $x$ of a closed surface determines uniquely its topology only if $x$ is two or an odd number (less than two). For $x=2$ the surface is a sphere and for odd values the surface is nonorientable. For each of the other possible values of $x$, nonpositive and even, there are two closed surfaces, one orientable and one nonorientable. Though the two graphs of a surface duality correspondence determine the surface characteristic, Theorem 1 provides no control over which of the two surfaces the maps might be.

In trying to find an analog of Theorem 1 for orientable maps, it is natural to use the notion of *oriented edge-correspondence* between two graphs, $G_V$ and $G_W$. That is a 1–1 correspondence between the edges of $G_V$ with orientations and the edges of $G_W$ with orientations such that if two edges with orientations correspond then the same edges with opposite orientations correspond.

For any $G_V$ and $G_W$ which are the graphs of dual oriented maps there is determined an oriented edge-correspondence with the following property. $P_0$: For each vertex $v$ of each graph, the edges directed away from $v$ correspond in the other graph to a directed subgraph which is connected and which has the same number of its edges directed away from as directed toward each of its vertices. Property $P_0$ follows from the fact that the faces of the oriented maps determine a *directed Euler path* in each of these directed subgraphs.

However, an oriented edge-correspondence with property $P_0$ between two connected graphs is not necessarily an oriented surface duality correspondence. That is we cannot always find directed Euler paths which fit together into single cyclic orders at each vertex. In particular, there are lots of connected graphs $G_V$, even where $V_V-E_V$ is odd, which are not graphs of oriented maps with one face, even though for any connected graph one can set up an oriented edge-correspondence satisfying $P_0$ with a graph which consists of one vertex and the proper number of edges. The directed-graph analog of Euler's theorem is true, but not true with enough leeway for the corresponding oriented surface-duality theorem to be true.

(Paper 69B1–142)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Maximum Matching and a Polyhedron With 0,1-Vertices[1]

**Jack Edmonds**

(December 1, 1964)

A matching in a graph $G$ is a subset of edges in $G$ such that no two meet the same node in $G$. The convex polyhedron $C$ is characterized, where the extreme points of $C$ correspond to the matchings in $G$. Where each edge of $G$ carries a real numerical weight, an efficient algorithm is described for finding a matching in $G$ with maximum weight-sum.

## Section 1

An algorithm is described for optimally pairing a finite set of objects. That is, given a real numerical weight for each unordered pair of objects in a set $Y$, to select a family of mutually disjoint pairs the sum of whose weights is maximum. The well-known optimum assignment problem [5][2] is the special case where $Y$ partitions into two sets $A$ and $B$ such that pairs contained in $A$ and pairs contained in $B$ are not positively weighted and therefore are superfluous to the problem. For this "bipartite" case the algorithm becomes a variant of the Hungarian method [3].

The problem is treated in terms of a graph $G$ whose nodes (vertices) are the objects $Y$ and whose edges are pairs of objects, including at least all of the positively weighted pairs. A *matching* in $G$ is a subset of its edges such that no two meet the same node in in $G$. The problem is to find a maximum-weight-sum matching in $G$. The special case where all the positive weights are one is treated in detail in [2] and [6]. The description here of the more general algorithm uses the terminology set up in [2]. Paper [2] (especially sec. 5) helps also to motivate this paper, though it is not really a prerequisate till section 7 here.

The increase in difficulty of the maximum weight-sum matching algorithm relative to the size of the graph is not exponential, and only moderately algebraic. The algorithm does not involve any "blind-alley programming"—which, essentially, amounts to testing a great many combinations.

The emphasis in this paper is on relating the matching problem to the theory of continuous linear inequalities. In particular, we prove a theorem analogous to one of G. Birkhoff [1] and J. von Neuman [5] which says that the extreme points of the convex set of doubly stochastic matrices (order $n$ by $n$) are the permutation matrices (order $n$ by $n$). That theorem and the Hungarian method are based on Konig's theorem about matchings in bipartite graphs. Our work is related to results on graphs due to Tutte [4].

There is an extensive related literature besides these references. One may refer to surveys on graphs, linear programming, network flow, and combinatorial analysis (other than enumeration). However, paper [2] and this one are together self-contained. For the algorithm without the polyhedral geometry, sections 4 and 7, here and in [2], suffice.

The technique, described in sections 2 and 3, of using linear programming duality to derive a description of the convex polyhedra associated with a class of combinatorial structures appears applicable, whereever the combinatorics is adequately understood, independently of the particular nature of the associated algorithm. The results of this paper suggest that, in applying linear programming to a combinatorial problem, the number of relevant inequalities is not important but their combinatorial structure is.

In another paper, I will extend the present work to "Optimum degree-constrained subgraphs". See the end of this paper for two main results of that extension.

## Section 2

Let the real variables $x_e \in E$ correspond to the edges $e$ of a finite graph $G$. Let $C$ be the convex polyhedron of vectors $< x >$ formed by the intersection of all the half-spaces given by the following inequalities (1), (2), and (3).

[1] Prepared while the author was a research associate at Princeton University, under the Princeton Logistic Research Project (Office of Naval Research). Based on work with the NBS Combinatorial Mathematics Project (Army Research Office, Durham).

[2] Figures in brackets indicate the literature references at the end of this paper.

(1) $x \geq 0$, for all $x \epsilon E$.

(2) for every node $v$ of $G$, $\Sigma x \leq 1$ (summed over $x \epsilon V$), where $V$ is the set of variables corresponding to the edges of $G$ which meet node $v$.

(3) for every subset $S$ of $2r+1$ nodes in $G$ where $r$ is a strictly positive integer, $\Sigma x \leq r$ (summed over $x \epsilon R$), where $R$ is the set of variables corresponding to the edges of $G$ with both ends in $S$.

Condition (I) is that the variables $x$ take on only values zero and one. Assuming condition (I), each vector $< x >$ represents and may be regarded as equivalent to the subset of edges in $G$ which correspond to the one-valued components of $< x >$.

Assuming (I), condition (2) is the definition of a matching in $G$. Assuming (I), condition (3) says that for any set $S$ of $2r+1$ nodes ($r$, a positive integer) the set $< x >$ of edges contains no more than $r$ edges with both endpoints in $S$; clearly this is implied by the set of edges being a matching in $G$. Therefore, assuming (I), condition (2) implies condition (3) for vectors $< x >$. However, replacing (I) by the weaker condition (1), it is easy to show that, where $G$ contains a circuit with an odd number of edges, condition (2) does not imply condition (3) for vectors $< x >$.

The essence of our following theorem ($P$) is that for purposes of linear-extremizing over the family of matchings in $G$, discreteness condition (I) and condition (2) can be replaced by the inequalities (1), (2), and (3).

Let $P$ be the set of vectors $< x >$ such that each component is a zero or one and such that the one-components correspond to the edges of a matching in $G$. That is let $P$ be the vectors $< x >$ which satisfy (I) and (2). We call $< x > \epsilon P$ a matching vector of $G$.

THEOREM (P): P *is the set of vertices (extreme points) of polyhedron* C.

Unless the graph $G$ has an edge joining each pair of nodes, inequalities (1), (2), and (3) generally include more than the minimal set of bounding planes for $C$, but that is not so important. What is important in order to provide a good characterization for maximum weight-sum matchings in a graph $G$ is a good characterization of *some* family of inequalities which together bound precisely the convex hull of $P$.

It is obvious that the points $P$ are vertices of $C$—that is they belong to $C$ and none lie half way between two other points of $C$. In fact, they are vertices of the larger polyhedron, $C'$, given by inequalities (1) and (2), and they are not sliced away by (3). However, in $C'$ there are other vertices. What remains to be proved is that vertices $P$ are the *only* vertices in $C$.

Every linear form in the variables of the space of a convex polyhedron is maximized over points in the polyhedron, if a maximum exists, by a vertex (and perhaps other points as well). Conversely, every vertex is the unique maximum over the polyhedron of some linear form.

In particular, where each edge $e$ of $G$ carries a real weight $c$, theorem ($P$) implies that the maximum weight-sum for matching in $G$ equals the maximum of

(4) $W = \Sigma cx$ (summed over $x \epsilon E$),

where real vector $< x >$ satisfies (1), (2), and (3). And conversely, theorem ($P$) follows by proving that for all real $< c >$, $W$ is maximized by a vector $< x >$ whose components are zeroes and ones.

## Section 3

Where $[a_{ij}]$ is any real matrix and $< b_j >$ and $< c_i >$ are real vectors, let $< x_i >$ represent the vectors satisfying the inequalities $x_i \geq 0$ *and* $\Sigma a_{ij} x_i \leq b_j$ (summed over $i$) and let $< y_j >$ represent the vectors satisfying the inequalities $y_j \geq 0$ and $\Sigma a_{ij} y_j \geq c_i$ (summed over $j$). The duality theorem of linear programming says that $\max \Sigma c_i x_i = \min \Sigma b_j y_j$ when these extrema exist for vectors $< x_i >$ and $< y_i >$.

To get the linear program which is dual to maximizing $W$ of (4) in the polyhedron $C$, we introduce a new variable corresponding to each inequality of (2) and (3). That is for each node $v$ in $G$ we introduce a variable, call it $y$, and for each set $S$ of $2r+1$ nodes in $G$ ($r$, a positive integer) we introduce a variable, call it $z$.

Let $< y, z >$ denote the vector of all variables $y$ and $z$. The duality theorem says that the maximum of $W$ is equal to the minimum of

(5) $U = \Sigma y + \Sigma rz$ (first sum taken over all nodes $v$ and second sum taken over all sets $S$), where vector $< y, z >$ satisfies the following nonnegativity inequalities (6), and the following inequalities (7), obtained from the transpose of the matrix of coefficients of inequalities (2) and (3).

(6) For every node $v$, $y \geq 0$; for every set $S$, $z \geq 0$.

(7) For every edge $e$ of $G$ and the nodes $v_1$ and $v_2$ which it meets, $y_1 + y_2 + \Sigma z \geq c$ (where $y_1$ and $y_2$ are the variables corresponding to $v_1$ and $v_2$ and where the sum is taken over all $z$ for which the corresponding set $S$ contains $v_1$ and $v_2$). The $c$ is the coefficient in linear form $W$ of the $x$ which corresponds to edge $e$. In other words, $c$ is the weight on edge $e$.

For a fair sized graph $G$, vector $< y, z >$ has a huge number of components. However, in general any vector which is a vertex of a dual polyhedron has no more nonzero components than the total number of variables in the primal linear program. In particular, any vector $< y, z >$ with which we deal will have no more nonzero components than the number of edges in $G$.

As easily shown in general for dual linear programs, $W \leq U$ for all admissible vectors $< x >$ and $< y, z >$. Therefore to prove that any linear form $W$ is maximized in $C$ by some matching vector, that is to prove theorem ($P$), it is sufficient to display a vector $< x > \epsilon P$ and some vector $< y, z >$ satisfying (6) and (7) such that $W = U$. Notice that we are using here only the easy part of the duality theorem. In fact, by displaying the vectors as described we will be proving, for the class of programs given by $C$ and $W$, the harder part of the duality theorem, though that is incidental.

Conversely, theorem ($P$) and the duality theorem (including the harder part) imply that a matching $M$ in $G$ has maximum weight-sum if and only if that weight-sum equals $U$ for some $< y, z >$ (with a mod-

erate number of nonzero components) which satisfies (6) and (7).

For any coefficients $< c >$ for $W$, that is for any weights $< c >$ on the edges of $G$, we shall describe a matching vector $< x >_0 \epsilon P$, corresponding to a matching $M$ in $G$, and a vector $< y, z > = < y, z >_0$ such that (6) and (7) hold and furthermore such that the following (8), (9), and (10) hold.

(8) $y = 0$ for each node $v$ which is not an endpoint of an edge in $M$.

(9) equality holds for each instance of (7) where $e \epsilon M$.

(10) for each $z > 0$, the set $S$ contains both endpoints of $r$ edges in $M$.

Therefore, summing together on each side the equations of (8) and (9), we will get an instance of $U = W$, where $< y, z > = < y, z >_0$ and $< x > = < x >_0$.

## Section 4

We turn now to theorem (*M*) which, for any matching $M$ corresponding to a matching vector $< x >_0$ which maximizes $W = cx$, will yield a vector $< y, z >_0$, satisfying (6), (7), (8), (9), and (10). From this and from the discussion in section 3, it follows that the existence of a vector $< y, z >_0$ is a necessary and sufficient condition for a matching $M$ to have maximum weight-sum. Theorem (*M*) itself is also such a condition in a different form. Theorem (*M*) displays the existence of much tighter and more complex structure than simply the vector $< y, z >_0$, tighter and more complex than really necessary for characterizing maximum matchings.

Ellis Johnson and Charles Zahn in studying this theory found that theorem (*M*) can be bypassed in obtaining the vector $< y, z >_0$ and that the maximum matching algorithm can be executed in terms of parameters of type $y$ and $z$ rather than the numerical parameters of theorem (*M*). One type of parameter may be arithmetically more convenient than the other, though the same combinatorial manipulations seem essential in either case. Theorem (*M*) seems justified by the insight it provides and by its natural relationship to the combinatorial manipulations of the algorithm, so it might as well be proved on the way to proving theorem (*P*). It is a direct consequence of the algorithm as described in section 7. More important, it is part of the description.

THEOREM (M): *For graph* G *with edge-weights* $< c >$, *a matching* M *is maximum if and only if there exists a sequence* $\{G_i\}$ $i = 0, \ldots, n$.

Each $G_i$ is a graph together with a matching $M_i$, a numerical weight $w(e^i)$ for each edge $e^i \epsilon G_i$, and a numerical weight $w(v^i)$ for each node $v^i \epsilon G_i$. Sequence $\{G_i\}$ has the following properties.

(a) $G_0$ is graph $G$ with edge-weights $w(e) = c$, with matching $M_0 = M$, and with any vertex-weights satisfying the general conditions below.

(b) $w(v^i) \geq 0$ for all $v^i \epsilon G_i$ and $w(v_1^i) + w(v_2^i) \geq w(e^i)$ for all $e^i \epsilon G_i$ where $v_1^i$ and $v_2^i$ are the endpoints of $e^i$.

(c) For $i = 0, \ldots, n-1$, there is in $G_i$ a circuit (simple closed path) $B_i$ containing $2a_i + 1$ edges, $a_i$ of them in $M_i$. Circuit $B_i$ is called a *blossom* in $(G_i, M_i)$.

(d) $w(v_1^i) + w(v_2^i) = w(e^i)$ for $e^i \epsilon B_i$.

(e) If a vertex of $B_i$ meets no edge of $M_i$, it will be one of the vertices, say $q^i$, with smallest weight, $w(q^i)$, in $B_i$.

(f) To obtain graph $G_{i+1}$ from graph $G_i$ shrink $B_i$ and all edges with both end-points in $B_i$ to a single node $u^{i+1}$ in $G_{i+1}$ which, in place of the vertices of $B_i$, is the new endpoint of those edges having with respect to $G_i$ one endpoint in $B_i$. The matching in $G_{i+1}$ is $M_{i+1} = M_i \cap G_{i+1}$.

(g) Weights in $G_{i+1}$ are the same as corresponding weights in $G_i$ except at $u^{i+1}$ and edges meeting $u^{i+1}$.

(h) For minimum $w(q^i)$ in $B_i$, $w(u^{i+1}) \leq w(q^i)$.

(i) For each edge $e^{i+1}$ meeting $u^{i+1}$, let $e^i$ be the corresponding edge in $G_i$, meeting $v^i$ of $B_i$. Then

$$w(e^{i+1}) = w(e^i) - w(v^i) + w(q^i).$$

(j) For $G_n$, $w(v_1^n) + w(v_2^n) = w(e^n)$ for all $e^n \epsilon M_n$.

(k) For a vertex $v^n \epsilon G_n$ not meeting an edge in $M_n$, $w(v^n) = 0$.

## Section 5

Vector $< y, z >_0$ is as follows. The node-sets $S$ for which the corresponding $z$ in $< y, z >_0$ is positive are among those, say $S_i$, corresponding to the $B_i$ of theorem (*M*). The nodes in each $S_i$ are the nodes of $G$ which have been absorbed into the nodes of $B_i$ (i.e., into $u^{i+1}$) in the process of going from graph $G = G_0$ to graph $G_i$. Define

(11) $d_i = w(q^i) - w(u^{i+1})$ for each $S_i$, and $d = 0$ for every other set $S$.

(12) $z = 2d$, for all $S$.

For each $y$ component of $< y, z >_0$, corresponding to node $v$ of $G$, we set

(13) $y = w(v) - \Sigma d$ (summed over $d$ corresponding to sets $S$ for which $v \epsilon S$), where $w(v)$ is the weight assigned by theorem (*M*) to node $v$ in $G_0 = G$.

PROOF: For each edge $e = e^0$ with endpoints $v_1$ and $v_2$ in $G = G_0$, let $e^j$ with endpoints $v_1^j$ and $v_2^j$ be the corresponding edge in $G^j$, where $j$ is such that either $j = n$ or else $e^j$ has both ends in $B_j$ and thus is absorbed into $u^{j+1}$.

By virtue of (g), in the applications of formula (i) to $e^j$ and its pre-images, each $w(v^j)$ term is either $w(v_1)$ or $w(v_2)$ or else a certain $w(u^{k+1})$ where $u^{k+1}$, the contraction of a $B_k$, absorbs one but not both of $v_1$ and $v_2$ in going from $G_0$ to $G_n$. Conversely, each such $w(u^{k+1})$ is either $w(v_1^j)$ or $w(v_2^j)$ or else a certain $w(v^j)$ term in one of the applications of (i) to $e^j$ and its pre-images. Furthermore, $w(v_1)$ is a certain $w(v^j)$ term if and only if $w(v_1^j)$ is a certain $w(u^{k+1})$. Otherwise, $w(v_1) = w(v_1^j)$. Similarly for $w(v_2)$. Thus by virtue of (g), repeated application of (i) and then repeated substitution of (11), yields:

(14) $w(e^j) = w(e) + w(v_1^j) + w(v_2^j) - w(v_1) - w(v_2) + \Sigma d_k$, where $k$ is summed over all the $B_k$ (or $u^{k+1}$) into which either $v_1$ or $v_2$, but not both, are absorbed in going from $G_0$ to $G_n$.



757-615 O-65-9

By (b) and (14),

(15) $w(v_1)+w(v_2)-\Sigma d_k \geq w(e)=c$, where $k$ is summed as in (14). Substituting (13) twice and (12) several times we get (7).

By (d), (j), and (14), equality holds for (15) and thus for (7) when $e \epsilon M$. Hence (9) is verified.

Substituting (11) for nonzero $d$ in (13), recollecting terms, and using (g), we have

(16) $y=w(v^n)+\Sigma[w(v^k)-w(q^k)]$, where $k$ is summed over all the $B_k$ into which $v$ is absorbed and where $v^k$ and $v^n$ are the images of $v$ in $B_k$ and $G_n$. Hence, by (b) and (e), we have $y \geq 0$. By (h) and (11), $z \geq 0$. Thus, (6) is verified.

When a node $v$ meets no edge of $M$, by (c) and (f) neither does any image $v^i$ meet an edge of $M_i$. Therefore by (e) and (k) and (16), $y=0$ in this case. Thus, (8) is verified.

Condition (10) follows from (c) and (f). Thus, modulo the "only if" part of theorem ($M$), conditions (6) through (10) are verified for some $< y, z >_0$ and theorem ($P$) is proved.

## Section 6

Since $U \geq W$ in general, the "if" part of theorem ($M$) is proved by the above translation from $\{G_i\}$ to a $< y, z >_0$ for which $U=W$.

It is much simpler for the matching algorithm, given $< c >$, to construct *some* maximum matching with a $\{G_i\}$ than to construct a $\{G_i\}$ for a particular maximum matching. Indeed, the simpler construction is all that is needed to get theorem ($P$). However, by a continuity argument we can show the existence of a $\{G_i\}$ for any maximum matching $M$ and thereby complete the proof of theorem ($M$).

Add a positive $\epsilon$ to the weight $c$ of each edge $e$ in $M$; leave all other edge weights the same. With this new $< c >$, $M$ is the unique maximum matching and therefore, by the simpler construction, the algorithm in section 7, there exists a $\{G_i\}$ for it.

An infinite sequence of sequences $\{G_i\}$, corresponding to an infinite sequence of $\epsilon$'s which approach zero, contains only a finite number of different combinatorial configurations aside from the values of the weights. The space of all possible weights is bounded and finite-dimensional. Hence there is a subsequence of $\{G_i\}$'s, combinatorially the same, with a limit which is a $\{G_i\}$ like we want.

## Section 7

The maximum-weight-sum matching algorithm consists of the maximum cardinality algorithm in [2] (secs. 4 and 7) together with small modifications suggested by theorem ($M$).

Suppose we have a sequence $\{G_i\}$ $(i=0, \ldots, m)$, with not-necessarily-maximum matchings $\{M_i\}$, satisfying all the conditions (a) through (j), omitting only (k). To get a weak sequence of this kind to start with—take $m=0$, take the matching to be empty, and take sufficiently large vertex weights.

We apply the algorithm in [2] to matching $M_m$ in the subgraph, $G'_m$, of $G_m$ which consists of all nodes of $G_m$ plus all edges $e^m$ for which $w(v_1^m)+w(v_2^m)=w(e^m)$. If $G_m$ does not satisfy (k) then there is an exposed node $r$ in $G'_m$, for which $w(r)>0$. (*Exposed* means it meets no matching edge.) Start growing in $G'_m$ a planted tree rooted at $r$.

If it grows into a flowered tree with blossom $B_m$, then in $G_m$ shrink $B_m$ to a $u^{m+1}$ to obtain a $G_{m+1}$. Where $q_m$ is the node in $B_m$ with smallest weight, set $w(u^{m+1})$ equal to $w(q^m)$ and adjust the weights of the edges which meet $u^{m+1}$ according to the formula in (i). Leave other weights in $G_{m+1}$ the same as in $G_m$. Thus, weights in $G_{m+1}$ will satisfy conditions (b), (g), (h), (i), and (j). Furthermore, $G'_{m+1}$ is the image in $G_{m+1}$ of $G'_m$, and so we continue in $G'_{m+1}$ with the tree image.

Eventually, possibly after a number of shrinkings, we obtain either: (1) an augumenting tree or, (2) a tree which has an outer vertex $v$ with $w(v)=0$ or, (3) a Hungarian tree which has outer vertices all with positive $w$.

In case 1, augment. That is, interchange matching roles of edges in the augmenting path. This yields a matching with larger weight-sum in the graph, still call it $G_m$, and disposes of one or two vertices violating (k). In case 2, the path in the tree joining $v$ to $r$ is really "augmenting" also, though such paths are not encountered in [2]. Treat it like the path in case 1 and it serves the same purpose.

The fact that a matching is obtained which has larger weight-sum is not needed for the validation of the algorithm. The important fact is the decrease in the number of nodes $r$ in the final term $G_m$ of $\{G_i\}$, such that $r$ is exposed and $w(r)>0$. After an augmentation in case 1 or 2, abandon the tree and start a new tree at another node $r$ if there is one.

When we get, in case 1 and 2, a matching with larger weight-sum in $G_m$, it yields a matching with larger weight-sum in each graph $G_i$ back through $G_0$. It does this by the process of successively selecting a new matching within each blossom $B_i$ which is compatible in $G_i$ with the matching already chosen with respect to $G_{i+1}$. If $u^{i+1}$, the blossom $B_i$ before "expansion", is an exposed node in $G_{i+1}$ then the selection of matching edges in $B_i$ is determined by condition (e). Sequence $\{G_i\}$ with these new matchings $M_i$ satisfies (a) through (j) and comes closer by at least one to satisfying (k). There is no advantage in selecting the matching in graphs $G_i$ other than $G_m$ until the algorithm is otherwise finished or until it is necessitated by condition (h) in case (3).

It is important for the type of step of the algorithm necessitated by (h) in case 3, described below, that we be able to "expand" the shrunken blossoms $B_i$ in an order different from the order in which they were shrunk—thereby obtaining a sequence $\{G_i^a\}$ of type $\{G_i\}$ (possibly not satisfying (k)) such that the terms $G_0^a$ and $G_m^a$ together with their weights and matchings are, respectively, identical to the terms $G_0$ and $G_m$ of $\{G_i\}$ and such that the blossoms $B_i^a$ of $\{G_i^a\}$ correspond in a different order to the blossoms $B_i$ of $\{G_i\}$. The order of the blossoms $B_i$ is not arbitrary but it is limited only by the relation of one $u^{i+1}$ having been absorbed into another.

In particular, for any node $v^m$ of $G_m$ which is the image of some $B_i$, let $B_k$ be the blossom of $\{G_i\}$ whose image in $G_m$ is $v^m$ and such that $k$ is maximum. Node $v^m$ can be "expanded" to obtain from $G_m$ with its weights and matching a certain graph, $G_{mk}$, with weights and a matching and a blossom, $B_{mk}$, corresponding to $B_k$. The graph structure, the weights, and the matching of $G_{mk}$ in the neighborhood of $B_{mk}$ are determined by $G_m$ in the neighborhood of $v^m$ and by $G_k$ in the neighborhood of $B_k$, subject to the conditions (other than (k)) of theorem (*M*). Elsewhere $G_{mk}$ is like $G_m$. There exists a sequence $\{G_i^q\}$ as described above such that the next to the last term is $G_{mk}$. Without describing it and justifying it here in any greater detail, the derivation of $\{G_i^q\}$ should be fairly evident from the local nature of the changes in successive terms of any sequence $\{G_i\}$.

In implementing the algorithm it may be better to represent the partial order structure of inclusion for the blossoms $B_i$ of a $\{G_i\}$ so that no special preference is given to $\{G_i\}$ or any of the other sequences $\{G_i^q\}$. Here we describe the algorithm with respect to a particular $\{G_i\}$ merely for convenience of exposition.

In case 3, the weights of the vertices in the Hungarian tree are adjusted. Weights of outer vertices are lowered and weights of inner vertices are raised by a uniform amount, which is as large as possible without violating for the sequence $\{G_i\}$ either (b) by making an outer vertex too small or (h) by making an inner vertex which is the image of a shrunken blossom too large. In the adjustment, property $w(v_1^m)+w(v_2^m)=w(e^m)$ is preserved for edges of the tree. Edges of $G'_m$ meeting an inner vertex and not an outer vertex will drop out of $G'_m$.

If (b) limits the adjustment, it may be because some outer vertex weight becomes zero. Then, possibly after the operation performed in case 2, we again have one less node violate (k). Otherwise if (b) limits the adjustment, it is because for one or more edges $e^m$ in $G_m$ and not in $G'_m$, which meet outer vertices, we get $w(v_1^m)+w(v_2^m)=(e^m)$. These edges enter $G'_m$, so the tree is no longer Hungarian and can be grown some more.

If, in case 3, condition (h) limits the adjustment at an inner vertex, say $v^m$, of the tree, it is because, where $\{G_i^q\}$ is the modification of $\{G_i\}$ cited above, $w(v^m)$ becomes as large as the smallest node-weight in $B_{mk}$ of $G_{mk}$. By substituting $\{G_i^q\}$ for $\{G_i\}$, and calling it $\{G_i\}$, we may regard $v^m$ as $u^m$, the image of $B_{m-1}$ in $G_m$ of $\{G_i\}$.

Node-weight $w(u^m)$ was first set equal to the smallest node weight $w(q^{m-1})$ in $B_{m-1}$ of $G_{m-1}$. However, by some weight adjustment where $u^m$ was an outer vertex of a Hungarian tree, $w(u^m)$ may have gotten smaller for a while—before the current adjustment gave $w(u^m)=w(q^{m-1})$ again.

We now describe what to do when $u^m$, the image of $B_{m-1}$, is an inner vertex of the current tree and $w(u^m)=w(q^{m-1})$. The pre-image in $G_{m-1}$ of the tree-edges is not generally one tree in $G_{m-1}$ but two. However, as described in section 7 of [2], using only edges from $B_{m-1}$, this pre-image can always be completed to a planted tree, say $T$, in $G'_{m-1}$, where $G'_{m-1}$ is the subgraph of $G_{m-1}$ consisting of edges $e^{m-1}$ for which $w(v_1^{m-1})+w(v_2^{m-1})=w(e^{m-1})$. Thus, we abandon $G_m$, replace $\{G_i\}$ $(i=0, \ldots, m)$ by $\{G_i\}$ $(i=0, \ldots, m-1)$, and continue the algorithm with respect to $T$ in $G'_{m-1}$.

This completes the description of the situations and operations in the algorithm. After handling a number of impediments like described in case 3 and like blossoms to be shrunk, the tree will grow into one which allows a decrease in the number of nodes violating (k). Finally when (k) is no longer violated in the final term of $\{G_i\}$, we have a $\{G_i\}$ satisfying all the conditions of theorem (*M*). It is easy to verify that all the conditions (a) through (j) are preserved by the steps of the algorithm.

The progress of the algorithm is measured first according to the decrease in the number of node weights which violate condition (k). The algorithm consists, in the large, of "growing" trees, one after the other. Each tree is abandoned when and only when it yields a decrease of one or two in that number. Thus, at most $N$ trees are grown where $N$ is the number of nodes in $G$.

The progress in the growth of an individual tree can be measured according to the number of distinct edges which have entered the tree, including those which have disappeared into blossoms shrunken while growing the tree. The latter never reappear in the same tree because they are absorbed into outer vertices of the tree, whereas only inner vertices of the tree are expanded in the course of its growth. The total number edges which ever enter a tree is less than $N$, and each edge enters at most once. It can be shown by a survey of the arithmetic involved and the combinatorics involved as described in [2] that an ample upper bound on the order of work between additions of edges to a tree and in the transition from each tree is $N^2$. This is assuming that the work involved in the individual arithmetic additions and subtractions is fixed. The amount of work in the algorithm also increases some according to the number of significant decimal places in the edge-weights.

## Section 8

THEOREM (P) generalizes to the following: *All the extreme points of polyhedra* I *and* II *consist of integer components.* As before the variables $x \epsilon E$ correspond to the edges of a graph $G$. Now there is an integer "capacity", d, associated with each node $v \epsilon G$. Where all d = 1, both I and II are polyhedron C of theorem (P).

I. (1) $x \geq 0$, for all $x \epsilon E$.

(2) $\Sigma x \leq d$ $(x \epsilon V)$, for all $V$, where $V$ corresponds to the edges which meet a node $v$ of $G$ and where $d$ is the integer capacity assigned to $v$.

(3) $\Sigma x \leq r (x \epsilon R)$, for all $R$ and $r$, where $r$ is a positive integer, and where $R$ corresponds to the edges of $G$ with both ends in a set $S$ of nodes such that $\Sigma d = 2r+1$ ($d$ summed over capacities of nodes $v \epsilon S$).

II. (1) $0 \leq x \leq 1$, for all $x \epsilon E$.

(2) $\Sigma x \leq d$ $(x \epsilon V)$, for all $V$, same as in I.

(3) $\Sigma x \leq r$ $(x \epsilon R)$, for all $R$ and $r$, where $r$ is a posi-

tive integer, and where $R$ corresponds to all edges with both ends in a set $S$ plus a number $t$ of edges with one end in $S$, where less than $d$ of the $t$ edges meet node $v$ in $S$, and where $2r+1=t+\Sigma d$ ($d$ summed over capacities of nodes $v \epsilon S$).

Correspondingly, there is an efficient algorithm for finding maximum-weight-sum, degree-constrained subgraphs.

(Paper 69B1–143)

## References


[1] G. Birkhoff, Tres. observaciones sobre el algebra lineal, Rev. Univ. Nac. Tucuman, Ser. A, **5,** 147–148 (1946).

[2] J. Edmonds, Paths, trees, and flowers, to appear in the Canadian J. Math. (May 1965).

[3] H. W. Kuhn, Variants of the Hungarian method for assignment problems, Naval Res. Logist. Quart. **3,** 253–258 (1956).

[4] W. T. Tutte, A short proof of the factor theorem for finite graphs, Canadian J. Math. **6,** 347–352 (1954).

[5] J. Von Neuman, A certain zero-sum two-person game equivalent to the optimum assignment problem, in Contributions to the Theory of Games, II (Ann. Math. Studies **38,** 5–12, ed. Tucker and Kuhn), Princeton (1953).

[6] C. Witzgall and C. Zahn, Jr., A modification of Edmonds' maximum matching algorithm, J. Res. NBS 69B (Math. and Math. Phys.) No. 1,

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 69B, Nos. 1 and 2, January–June 1965

# Publications of the National Bureau of Standards*

## Selected Abstracts

**Simple derivation of the faxen solution to the Lamm equation,** G. H. Weiss, *J. Math, Phys.* **5,** *No. 5, 675–676 (May 1964).*
In this paper a Hankel transform technique is used to derive the Faxén solution to the Lamm equation when the sedimentation coefficient is constant and when it varies linearly with concentration.

**The basis of the functional assumption in the theory of the Boltzmann equation,** M. S. Green and R. A. Piccirelli, *Phys. Rev.* **132,** *No. 3, 1388 (Nov. 1963).*
The long-time behavior of the $n$-particle probability densities for a large, dilute system of point particles interacting with short-range forces is studied. The main result is an exact series for the $n$-particle density which consists of two parts. The first part is a time-independent functional of the singlet density which is expressed as a functional power series and which is a direct analog of the equilibrium density series. The second part is also a functional power series in the singlet density but the coefficients depend on time and on the initial correlations. The coefficients of both series are given explicitly in terms of operators which are determined by the dynamics of isolated groups of particles. It is demonstrated that these operators vanish for phase points corresponding to motions during which there are two or more groups of particles which either are statistically and dynamically independent or are such that each of them is dynamically connected to the rest by no more than one particle. It is argued that all the terms of the exact series are finite and that the terms of second part (the error) decrease with increasing time so that the first part is the asymptotic form proposed by Bogoliubov. The relevance of the results for the Boltzmann equation is indicated.

**Elementary derivation of time-correlation formulas for transport coefficients,** R. Zwanzig, *J. Chem. Phys.* **40,** *No. 9, 2527–2533 (May 1964).*
This article presents an elementary derivation of time-correlation formulas for transport coefficients. The derivation, based on classical statistical mechanics, is modelled on one by Kubo, Yokota, and Nakajima. In agreement with theirs, it makes use of the assumption of regression of fluctuations. It differs, however, in its avoidance of Markoffian assumptions. The decisive point of the derivation is the introduction of a parameter $\xi$ to measure the rate of the transport process. In the limit of a very slow process, i.e., in the limit $\xi \to 0$ and time $t \to \infty$ with $\xi t$ held constant, the results coincide with those obtained by Markoffian assumptions. The relation to other work, and in particular to Kirkwood's friction constant theory, is touched on briefly.

**A complete description of the normal subgroups of genus one of the modular group,** M. Newman, *Am. J. Math* **86,** *17–24 (1964).*
The normal subgroups of genus one of the modular group are put into 1–1 correspondence with three integer parameters $(p, m, d)$ satisfying

$$p > 0,\ 0 \le m \le d-1,\ m^2 + m + 1 \equiv 0 \pmod{d}.$$

**The zeros of infrapolynomials with prescribed values at given points,** J. L. Walsh and O. Shisha, *J. Math Anal. and Appl.* **7,** *No. 2, 176–182 (Oct. 1963).*
We consider infrapolynomials $A(z)$ for which the values of $A(z)$, $A'(z)$, . . ., $A^{(m)}(z)$ at given points $z_1, z_2, \ldots, z_k$ are prescribed. Some results regarding the geometric location of the zeros of such infrapolynomials are proved, including analogues of the theorems of Bôcher and Jensen.

**A symmetry group containing both the Lorentz group and $SU_3$,** A. O. Braut, *Il Nuovo Cimento* **32,** *234–236 (Apr. 1964).*
The purpose of this note is to present a group, based on analytic transformations of a Kaehler manifold, which contains the Lorentz group and accounts for all the quantum numbers of strong interactions, including baryon and lepton numbers, and no more.

**Error analysis of Miller's recurrence algorithm,** R. W. J. Olver, *Math. Comp.* **18,** *No. 85, 65–74 (Jan. 1964).*
Miller's algorithm is a device for computing the most rapidly decreasing solution of a second-order linear difference equation. In this paper strict upper bounds are given for the errors in the values yielded by the algorithm, and general conclusions are drawn concerning the accuracy of the process.

**The spheroidal method for satellite orbits,** J. P. Vinti, *Proc. Intern. Symp. Use of Artificial Satellites for Geodesy, p. 12–16 (North Holland Publ. Co., Amsterdam, The Netherlands, 1963).*
To calculate the orbit of an artificial satellite of an oblate planet, the author has devised a very accurate approximation for its gravitational potential, leading to separability of the problem in oblate spheroidal coordinates. The present paper sketches the procedures for finding this potential and for calculating the corresponding orbit. It then discusses the remaining perturbing potential and a possible method of applying the calculated perturbed orbit for accurate determination of the earth's gravitational field.

**Error bounds for asymptotic expansions in turning-point problems,** F. W. J. Olver, *Soc. Ind. Appl. Math.* **1,** *No. 1, 200–214 (Mar. 1964).*
Error bounds are derived and examined for approximate series solutions in terms of Airy functions of the differential equations

$$\frac{d^2w}{dx^2} = \{x + f(x)\}w, \quad \frac{d^2w}{dx^2} = \{u^2p(u, x) + q(u, x)\}w.$$

The coefficients $f(x)$, $p(u, x)$ and $q(u, x)$ are free from singularities and $p(u, x)$ has just one zero in the region of integration, which can be either an interval of the real axis or a domain in the complex plane and in neither case need be bounded. For large values of the parameter $u$ the series in question are uniform asymptotic expansions of solutions of the second differential equation.
An application is made to the Bessel functions $J_\nu(x)$ and $Y_\nu(x)$, and bounds are obtained for the remainder terms in the uniform asymptotic expansions of these functions for large values of $\nu$.

**On the identity of three generalized master equations,** R. Zwanzig, *Physica* **30,** *1109–1123 (1964).*
Three apparently different quantum mechanical master equations, derived by Prigogine and Résibois, by Montroll, and independently by Nakajima and Zwanzig, are shown to be identical. The derivation by Zwanzig, based on projection operator and Liouville operator techniques, is repeated in greater detail than in previous articles. The results of Prigogine and Résibois, and of Montroll, are found by making changes in notation.

**On the validity of the Einstein-Smoluchowski theory light scattering,** R. Zwanzig, *J. Am. Chem. Soc.* **86,** *3489–3493 (Mar. 30, 1964).*
The molecular basis of the Einstein–Smoluchowski macroscopic theory of light scattering is investigated, using a generalization of Frölich's theory of dielectrics. The macroscopic and molecular theories differ by an experimentally negligible quantity, associated with the electric field dependence of the dielectric constant. A slight depolarization, not accounted for by the macroscopic theory, is associated with the same quantity. Because of the method of

derivation, the results obtained here are valid only in the limit of long wave length and low frequency.

**Symplectic modulary groups,** M. Newman and J. R. Smart, *Acta Arithmetica* **IX,** *83–89 (1964).*
The principal results of this paper are (1) If a matrix $A$ is symmetric modulo $n$ then there is a symmetric matrix $B$ congruent to $A$ modulo $n$, and (2) If a matrix $A$ is symplectic modulo $n$ then there is a symplectic matrix $B$ congruent to $A$ modulo $n$. These are applied to the determination of the structure of the quotient groups $\Gamma(a)/\Gamma(b)$, $a|b$ where $\Gamma$ is the $2t \times 2t$ symplectic modular group and $\Gamma(n)$ the principal congruence subgroup of $\Gamma$ of level $n$.

## Other NBS Publications

**J. Res. NBS 68A (Phys. and Chem.), No. 6 (Nov.–Dec.), 70 cents.**

The 1962 $He^3$ scale of temperatures. I. New vapor pressure comparisons. S. G. Sydoriak and R. H. Sherman.
The 1962 $He^3$ scale of temperatures. II. Derivation. S. G. Sydoriak, T. R. Roberts, and R. H. Sherman.
The 1962 $He^3$ scale of temperatures. III. Evaluation and status. T. R. Roberts, R. H. Sherman, and S. G. Sydoriak.
The 1962 $He^3$ scale of temperatures. IV. Tables. R. H. Sherman, S. G. Sydoriak, and T. R. Roberts.
Absolute isotopic abundance ratio and the atomic weight of a reference sample of copper. W. R. Sheilds, T. J. Murphy, and E. L. Garner.
Absolute isotopic abundance ratio and the atomic weight of bromine. E. J. Catanzaro, T. J. Murphy, E. L. Garner, and W. R. Sheilds.
Optical properties of thin films on transparent surfaces by ellipsometry; internal reflection for film covered surfaces near the critical angle. E. Passaglia and R. R. Stromberg.
On the second-order transition of a rubber. E. A. DiMarzio.
Ionization constants and reactivity of isomers of eugenol. G. M. Brauer, H. Argentar, and G. Durany.
Near infrared spectrophotometric method for the determination of hydration numbers. R. A. Durst and J. K. Taylor.
Franck-Condon factors for the ionization of $H_2$, HD, and $D_2$. M. E. Wacks.
Calculation of the geometrical structure of some $AH_n$ molecules. M. Krauss.
Preparation and heat of formation of a magnesium oxysulfate. E. S. Newman.
Heat capacity of potassium borohydride ($KBH_4$) from 15 to 375 °K thermodynamic properties from 0 to 700 °K. G. T. Furukawa, M. L. Reilly, and J. H. Piccirelli.
Heat of combustion and heat of formation of aluminum carbide. R. C. King and G. T. Armstrong.
Some elastic compliances of single crystal rutile from 25 to 1000 °C. S. Spinner and J. B. Wachtman, Jr.
Energy dependence of the D-D reaction cross section at low energies. J. G. Brennan and J. J. Coyne.
Steady-state response of silicon radiation detectors of the diffused P-N junction type to x rays. I: Photovoltaic mode of operation. K. Scharf and J. H. Sparrow.
Intercomparison of high-energy x-ray intensity measurements. J. S. Pruitt and S. R. Domen.

**J. Res. NBS 69A (Phys. and Chem.), No. 1, (Jan.-Feb. 1965), 70 cents.**

Heat of decomposition of sodium and potassium chlorate. A. A. Gilliland and D. D. Wagman.
Application of precise heat-capacity data to the analysis of the temperature intervals of the NBS-1955 and the International Practical Temperature Scales in the region of 90 °K. G. T. Furukawa and M. L. Reilly.
Heat capacity and thermodynamic properties of beryllium aluminate (chrysoberyl), $BeO{\cdot}Al_2O_3$, from 16 to 380 °K. G. T. Furukawa and W. G. Saba.
Thermodynamics of the ternary system: water-sodium chloride-barium chloride at 25 °C. R. A. Robinson and V. E. Bowe
Compressibility of eleven inorganic materials. C. E. Weir.
Determination of the intermolecular entanglement coupling spacings in polyisoprene by viscosity measurements. L. J. Fetters.
Electrical conductivity of dilute solutions of "sea water" from 5 to 120 °C. C. G. Malmberg.
Preparation of a carbonate-free complex calcium aluminate. H. A. Berman.
Synthesis and thermal stability of bis(8-hydroxyquinoline)-Schiff base coordination polymers. E. Horowitz and T. P. Perros.
Atomic absorption spectrum of praseodymium (Pr I). R. Zalubas and M. Wilson.

**Radio Sci. J. Res. NBS/USNC-URSI, 68D, No. 12, (Dec. 1964), $1.00.**

Experiment on the constancy of the velocity of electromagnetic radiation. P. Beckmann and P. Mandics.
Measurement of the phase velocity of VLF propagation in the earth-ionosphere waveguide. F. K. Steele and C. J. Chilton.
Wave hop theory of long distance propagation of low-frequency radio waves. L. A. Berry.
Wave propagation in a compressible ionosphere. Part I. S. R. Seshadri.
Wave propagation in a compressible ionosphere. Part II. S. R. Seshadri.
Spatial properties of the amplitude fading of continuous HF radio waves. J. Ames.
Physical properties of the polar winter mesosphere obtained from low-frequency propagation and partial reflection studies. J. S. Belrose, L. R. Bodé, and L. W. Hewitt.
Reply to the "Remarks by Donald H. Menzel With Reference to Bailey's Comments on Solar Electric Fields." V. A. Bailey.

**Radio Sci. J. Res. NBS/USNC-URSI, 69D, No. 1, (Jan. 1965), $1.00.**

Some problems of ionospheric nonlinearities. D. H. Menzel.
Some nonlinear phenomena in the ionosphere. V. A. Bailey.
An experimental study of gyro interaction in the ionosphere, at oblique incidence. F. H. Hibberd.
On some nonlinear phenomena in the ionospheric plasma. P. Caldirola and O. De Barbieri.
Ionospheric cross modulation: a microscopic theory. D. Layzer and D. H. Menzel.
VLF noise bands observed by the Alouette I satellite. J. S. Belrose and R. E. Barrington.
Excitation of optical radiation by high power density radio beams. L. R. Megill.
Alteration of the electron density of the lower ionosphere with ground-based transmitters. P. P. Lombardini.
Collision effects in hydromagneto-ionic theory. H. K. Sen and A. A. Wyller.
Electromagnetic wave reflection from an oscillating, collision-free magneto-ionic medium. O. E. H. Rydbeck.
Nonlinear propagation of electromagnetic waves in magnetoplasmas. II. (An invited abstract) M. S. Sodha and C. J. Palumbo.

**Radio Sci. J. Res. NBS/USNC-URSI 69D, No. 2, (Feb. 1965), $1.00.**

Electromagnetic wave penetration of reentry plasma sheaths. M. P. Bachynski.
On the use of refractive index diagrams for source-excited anisotropic regions. L. B. Felsen.
Surface waves along a perfectly conducting plane covered with semi-infinite magneto-plasma. S. Adachi and Y. Mushiake.
On the point of emergence of a microwave beam entering a linearly graded plasma. A. L. Cullen.
Self and mutual admittances of waveguides radiating into plasma layers. J. Galejs.
Effect of electron collisions on the formulas of magneto-ionic theory. K. G. Budden.
Momentum transfer collisions in oxygen for thermal electrons. M. H. Mentzoni.
Experimental studies of perturbations in ionospheric plasma. L. H. Heisler.
Electromagnetic scattering by gyrotropic cylinders with axial magnetic fields. W. C. Y. Lee, L. Peters, Jr., and C. H. Walter.
Radar cross sections of plasma bodies at the plasma frequency. L. Peters, Jr.
Electroacoustic waves excited by a space vehicle in ionized atmosphere and its effect on radar return. Kun-Mu Chen.

Dicussion on basic equations with source terms in compressive plasmas. Kun-Mu Chen.

Scattering of electromagnetic and electroacoustic waves by a cylindrical object in a compressible plasma. J. R. Wait.

Multiphase periodic very-low-frequency emissions. N. Brice.

Radio studies of the high-latitude ionosphere during the solar eclipse of 20 July 1963. R. D. Hunsucker.

VLF and LF fields propagating near and into a rough sea. R. M. Lerner and J. Max.

Insulated and loaded loop antenna immersed in a conducting medium. R. H. Williams.

Capacitance of biconical antennas in magneto-ionic media; elliptic cone capacitance. V. P. Pyati and H. Weil.

Calculations of the bistatic scattering cross section of a sphere with an impedance boundary condition. J. R. Wait and C. M. Jackson.

Concept of differential reflectivity as applied to the reflection of beam-limited radiation by a convex body. A. Erteza, J. A. Doran, and D. H. Lenhert.

---

Standard x-ray diffraction powder patterns, H. E. Swanson, M. C. Morris, E. H. Evans, and L. Ulmer, NBS Mono. 25-Sec. 3, (July 31, 1964), 40 cents.

Tables of electron radial functions and tangents of phase shifts for light nuclei (Z=1 through 10), C. P. Bhalla, NBS Mono. 81, (Aug. 6, 1964), $3.25.

Project FIST. Fault isolation by semi-automatic techniques, G. Shapiro, O. B. Laug, G. J. Rogers, and P. M. Fulcomer, Jr., NBS Mono. 83, (Sept. 17, 1964), 55 cents.

A complete mode sum for LF, VLF, ELF terrestrial radio wave fields, J. R. Johler and L. A. Berry, NBS Mono. 78, (Oct. 1, 1964), 20 cents.

A study of lunar surface radio communication, L. E. Vogler, NBS Mono. 85. (Sept. 14, 1964), 70 cents.

United States standard for the colors of signal lights, F. C. Breckenridge, NBS Handb. 95, (Aug. 21, 1964), 25 cents.

Standard Reference Materials: Preparation of NBS copper-base spectrochemical standards, R. E. Michaelis, L. L. Wyman, and R. Flitsch, NBS Misc. Publ. 260–2, (Oct. 15, 1964), 35 cents.

Hydraulic Research in the United States 1964, Ed. H. K. Middleton, NBS Misc. Publ. 261, (Aug. 14, 1964), $1.25.

Water-repellent preservative non-pressure treatment for millwork, CS262–63 (Dec. 31, 1963), 10 cents.

Quarterly radio noise data March, April, May 1963, W. Q. Crichlow, R. T. Disney, and M. A. Jenkins, NBS Tech. Note 18–18, (July 25, 1964), 50 cents.

Quarterly radio noise data June, July, August, 1963, W. Q. Crichlow, R. T. Disney, and M. A. Jenkins, NBS Tech. Note 18–19 (Aug. 21, 1964), 50 cents.

The electrical properties of aluminum for cryogenic electromagnets, R. J. Corruccini, NBS Tech. Note 218 (Aug. 30, 1964), 30 cents.

Some modifications in methods of calibration of universal ratio sets, D. Ramaley, NBS Tech. Note 220 (Aug. 30, 1964), 25 cents.

Demagnetizing factors for oblate spheroids used in ferrimagnetic resonance measurements, L. B. Schmidt, W. E. Case, and R. D. Harrington, NBS Tech. Note 221 (Sept. 4, 1964), 20 cents.

Concerning the theory of radiation from a slotted conducting plane in a plasma environment, J. R. Wait, NBS Tech. Note 223, (Sept. 28, 1964), 20 cents.

Bibliography on the measurement of bulk resistivity of semiconductor materials for electron devices, J. C. French, NBS Tech. Note 232 (Oct. 21, 1964), 60 cents.

A helicopter battery service simulator, W. G. Eicke, Jr., NBS Tech. Note 244, (Aug. 28, 1964), 20 cents.

Factors influencing the design of original-document scanners for input to computers, NBS Tech. Note 245 (Aug. 19, 1964), 35 cents.

Survey of magnetic thin film materials, G. W. Reimherr, NBS Tech. Note 247, (Sept. 25, 1964), 40 cents.

Radiochemical analysis: Activation analysis, instrumentation, radiation techniques, and radioisotope techniques, Ed. J. R. DeVoe, NBS Tech. Note 248 (Aug. 21, 1964), 50 cents.

A nomogram for computing $\frac{a+jb}{c+jd}$ and a nomogram for computing $\left|\frac{a+jb}{c+jd}\right|$, H. S. Bowman, NBS Tech. Note 250 (Oct. 2, 1964), 15 cents.

Tooth fluorometer, A. F. Forziati, J. Kumpula, and J. J. Barone, J. Am. Dental Assoc. **67,** No. 5, 663–669 (Nov. 1963).

Thermodynamic representation of high-pressure vapor-liquid equilibria, J. M. Prausnitz, Chem. Eng. Sci. **18,** 613–620 (1963).

Nuclear magnetic relaxation of the impurity nucleus in dilute ferromagnetic alloys, R. L. Streever, Phys. Rev. **134,** No. 6A, A1612–A1617 (June 15, 1964).

NBS microcopy resolution test chart, B. H. Fouquet, Proc. Natl. Microfilm Association Meeting, San Francisco, Calif., Apr. 30–May 13, 1963, pp. 67–74 (1963).

Metastable transitions in mass spectra of methane and the deuteromethanes, V. H. Diebeler and H. M. Rosenstock, J. Chem. Phys. **39,** 1326 (1963).

Electron distribution in the ionosphere, J. W. Wright, Proc. URSI XIVth General Assembly, Sept. 1963, Tokyo, Japan **XIII,** 206–222 (1964).

Studies of photodissociation of molecular ions, G. H. Dunn, Book, Atomic Collision Processes, pp. 997–1005 (North-Holland Publ. Co., Amsterdam, The Netherlands, 1964).

Studies at NBS of the platinum-6% rhodium vs. platinum-30% rhodium thermocouple—A preliminary report, G. W. Burns, Proc. SAE meeting, Sept. 23–27, 1963, Los Angeles, Calif., Preprint 750B, (1963).

Spread-*F* observations by the Alouette topside sounder satellite, W. Calvert and C. W. Schmid, J. Geophys. Res. **69,** No. 9, 1839–1852 (May 1, 1964).

Spectral distribution of typical daylight as a function of correlated color temperature, D. B. Judd, D. L. MacAdam, and G. Wyszecki, J. Opt. Soc. Am. **54,** No. 8, 1031–1040 (Aug. 1964).

Medical applications of refrigeration, E. F. G. Brentari and R. V. Smith, ASHRAE Guide and Data Book, ch. 68, 773–778 (Am. Soc. Heating, Refrig. Air Conditioning Engrs., Inc., New York, N.Y., 1964).

Mechanical and dielectric relaxation of crystalline polymers in relation to degree of crystallinity and morphology, E. Passaglia, Soc. Plastics Engrs., pp. 169–177 (July 1964).

Measuring the directivity of a directional coupler using a sliding short-circuit and an adjustable sliding termination, R. W. Beatty, IEEE Trans. Microwave Theory Tech. **MTT-12,** No. 3, 283 (May 1964).

Mass spectrometric study of the photoionization of acetylene and acetylene-$d_2$, V. H. Dibeler and R. M. Reese, J. Chem. Phys. **40,** No. 7, 2034–2035 (Apr. 1, 1964).

Magnetically retained evaporation mask, F. E. Jones and A. B. Castle, Sr., Rev. Sci. Instr. **34,** No. 9, 1055 (Sept. 1963).

Low-temperature infrared studies of the chemistry of free radicals, M. E. Jacox and D. E. Milligan, Appl. Opt. **3,** 873 (July 1964).

Characteristics energy losses of electrons in solids, L. Marton, Book, Actions Chimiques et Biologiques Des Radiations, ch. IV, pp. 260–281 (Masson et Cie Editeurs, Paris, France, 1964).

Isotopic fractionation of uranium related to roll features in sandstone, Shirley Basin, Wyoming, J. N. Rosholt, E. N. Harshman, W. R. Shields, and E. L. Garner, Econ. Geology **59,** 570 (1964).

Book, Thermal Degradation of Organic Polymers, S. L. Madorsky, Interscience Publ. Inc., New York, N.Y. (1964).

Solid state detector for beta-ray spectroscopy below 4.2°, K. J. Chin, A. T. Hirshfeld, and D. D. Hoppes, Rev. Sci. Instr. **34,** No. 12, 1258–1259 (Nov. 1963).

Should the convention definition of mismatch loss be abandoned? K. A. Norton, Proc. **IEEE 52,** 710 (June 1964).

Frequency and time standards, G. E. Hudson, Electronics World **72** 20–23 (Aug. 1964).

Recognition of clauses in machine translation of languages, F. L. Alt, Proc. Intern. Conf. Machine Translation of Languages and Applied Language Analysis, Teddington, England Symp. No. 13, pp. 126–142 (Her Majesty's Stationery Office, London, England, Sept. 1961).

X-ray study of a cadmium borate glass, S. Block, J. Molin, and J. Piermarini, J. Am. Ceram. Soc. **46,** 557 (1963).

World days program, A. H. Shapley, IQSY Notes **5,** 27–30 (Dec. 1963); **6,** 27–32 (Mar. 1964); **7,** 13–15, (May 1964).

Weights and measures and aerosols, M. W. Jensen, Proc. 50th Annual Meeting of Chemical Specialties Manufactures Association, pp. 33–35 (Dec. 9–12, 1963).

Unitary symmetry in photoproduction and other electromagnetic interactions, C. A. Levinson, H. J. Lipkin, and S. Meshkov, Phys. Letters **7,** No. 1, 81–84 (Oct. 15, 1963).

Transequatorial reception of a very-low-frequency transmission, C. J. Chilton, A. Diede, and S. Radicella, J. Geophys. Res. **69,** No. 7, 1319–1328 (Apr. 1, 1964).

Tomorrow's facilities for standards and measurement, H. F. Wollin, Scale J. 50, No. 6, 2 (Mar. 1964).

Review of recent developments of high temperature thermocouples, P. D. Freeze, Am. Soc. Mech. Engrs., Paper No. 63–Wa **212,** 16 (Nov. 1963).

Effects of magnetic disturbance as noted on oblique incidence ionograms, V. Agy and K. Davies, AGARDograph **59,** The effect of disturbances of solar origin on communications, Ed. G. J. Gassmann. pp. 167–176 (Pergamon Press, Inc., London, England, 1963).

Editorial, R. W. Beatty, IEEE Trans. Microwave Theory Tech. **MTT-12,** No. 3, 274 (May 1964).

Comparison calibration of inductive voltage dividers, R. V. Lisle and T. L. Zapf, ISA 1963 Proc. Paper 7.1.63 (1963).

Correction to a method for measuring the directivity of directional couplers, G. E. Schafer and R. W. Beatty, IEEE Trans. Microwave Theory Tech. **MTT-12,** No. 3, 361 (May 1964).

Ionospheric absorption in conjugate regions and possible oscillation of the ionosphere, H. J. A. Chivers and J. K. Hargreaves, Nature **202,** 891–893 (May 30, 1964).

Inline waveguide attenuator, W. Larson, IEEE Trans. Microwave Theory Tech. Corres. Sec. **MTT-12,** No. 3, 367–368 (May 1964).

Initial preparation and metastable transitions in mass spectra, H. M. Rosenstock, V. H. Dibeler, and F. N. Harllee, J. Chem. Phys. **40,** No. 2, 591–594 (Jan. 15, 1964).

Infrared spectroscopy of weak charge-transfer absorption spectra of NO in Kr and $CH_3OH$ solutions, E. E. Ferguson and H. P. Broida, J. Chem. Phys. **40,** 3715 (1964).

Index to Journal of Association for Computing Machines, W. W. Youden, J. Assoc. Comp. Mach. **10,** No. 4, 583–646 (Oct. 1963).

Inelastic electron scattering from rare gases, determination of oscillator strengths in the continuum, C. E. Kuyatt and J. A. Simpson, Proc. Third Intern. Conf. Physics of Electronic and Atomic Collisions, London, England, 1963, pp. 191–200 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

Improved transducer for external recording of arterial pulse waves, M. Davis, B. Gilmore, and E. Freis, IEEE Trans. Bio-Med. Elec., **BME-10,** No. 4, 173–175 (Oct. 1963).

Physical properties of cements, based on zinc oxide hydrogenated rosin, *o*-ethoxybenzoic acid and eugenol, S. Civijan and G. M. Brauer, J. Dental Res. **43,** No. 2, 281–299 (Mar.–Apr. 1964).

A study of geomagnetic effects associated with auroral zone electron precipitation observed by balloons, W. H. Campbell, J. Geomagnetism and Geoelectricity **11,** No. 1, 37–57 (June 1964).

The characterization of large single crystals by highvoltage X-ray Laue photographs, H. S. Peiser and E. P. Levine, (Proc. 11th Ann. Conf. on Applications of X-ray Analysis, Aug. 1962) Book, Advances in X-ray Analysis **6,** 158–163 (Plenum Press, Inc., New York, N.Y., 1963).

The calculations of autoionization probabilities, J. W. Cooper, Proc. Third Intern. Conf. Physics of Electronic and Atomic Collisions, London, 1963, pp. 595–599 (North Holland Publ. Co., Amsterdam, The Netherlands 1964).

On the constitution of the topside ionosphere, R. W. Knecht, H. Rishbeth, and T. E. VanZandt, (Proc. of NATO Advanced Study Inst., April 17–26, 1963, Skeikampen, Norway), Book, Electron Density Distribution in Ionosphere and Exosphere, pp. 348–350 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

On a problem of G. Sansone, M. Newman, Ann. Mat. **LXV,** 27–34 (1964).

A time-varying model of the ionospheric *F*2 layer, H. Rishbeth, J. Atmospheric Terrest. Phys. **26,** No. 6, 657–685 (June 1964).

The anomaly in the response of semiconductor detectors at low temperatures, W. R. Dodge, S. Domen, D. Hopper, and A. Hirshfeld, IEEE Trans. Nucl. Sci. **NS–11,** 1–14, (June 1964).

A theoretical study of the Martian ionosphere, R. B. Norton, Book, Advances in the Astronautical Science Exploration of Mars 15, 533–542 (Western Periodicals, California, 1963).

A new high voltage calibration facility at the National Bureau of Standards, Boulder, W. W. Scott, ISA 1963 Proc. Paper No. 7.2.63 (1963).

Energy dependence for the photodetachment of $I^-$ near threshold, B. Steiner, M. L. Seman, and L. M. Branscomb, Proc. Third Intern. Conf. Physics of Electronic & Atomic Collision, London, England, 1963, pp. 537–542 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

A low input VSWR coaxial diode switch for the UHF band, W. L. Ecklund, IEEE Trans. Microwave Theory Tech. Corres. Sec. **MTT-12,** No. 3, 359 (May 1964).

Suggested practices for establishing sampling and sample preparation techniques in spectrochemical analysis, R. E. Michaelis, Book, Methods for Emission of Spectrochemical analysis, 4th Edition, pp. 256–259 (American Society for Testing Materials, Phila., Pa. Feb. 1964).

S-wave hyperon-nucleon interactions and $SU_3$ symmetry, P. D. DeSouza, G. A. Snow, and S. Meshkov, Phys. Rev. **135,** 565 (1964).

Synthesis of higher ketoses by aldol reactions, II. Unsubstituted heptuloses, R. Schaffer, J. Org. Chem. **29,** 1471–1475 (June 1964).

Excess entropy at glass transformation, A. B. Bestul and S. S. Chang, J. Chem. Phys. **40,** 3731 (1964).

The large 50 Mc/s dipole array at Jicamarca Radio Observatory, G. R. Ochs, Digest of the 1963 IEEE Professional Group on Antennas and Propagation, p. 237 (July 1963).

The ionosphere and radio propagation in the lower frequencies (a report of an AGU session by the same title), T. N. Gautier Science **143,** No. 3611, 1197–1199 (Mar. 13, 1964).

Numerical analysis of the thermal environment of occupied underground spaces with finite cover using a digital computer, T. Kusuda and P. R. Achenbach, ASHRAE Trans. **6,** No. 6, 99–105 (June 1964).

Nuclear magnetic resonances of $^{69}Ga$ and $^{71}Ga$ in gallium-substituted yttrium iron garnet, R. L. Streever and G. A. Uriano, Phys. Rev. Letters **12,** No. 22, J516 1–3–J516 3–3 (June 1, 1964).

A diffusive model for the initial phase of a solar proton event, G. C. Reid, J. Geophys. Res. **69,** 2659–2667 (June 1964).

Emission spectrometry, B. F. Scribner and M. Margoshes, Anal. Chem. **36,** No. 5, 329–343R (1964).

Abbreviated Calendar record, January-February 1964, A. H. Shapley, IQSY Notes **7,** 19–25 (May 1964).

Evidence for a cooperative intramolecular transition in poly-L-proline, F. Gornick, L. Mandelkern, A. F. Diorio, and D. E. Roberts, Am. Chem. Soc. **86,** No. 13, 2549–2555 (July 1964).

The isomerization of D-manno-3-heptulose by alkali, R. Schaffer, J. Org. Chem. **29,** 1471–1475 (June 1964).

Insertion loss concepts, R. W. Beatty, Proc. IEEE **52,** 663–671 (June 1964).

Reaction rates for the formation of $He_2+$ and $Ne_2+$, E. C. Beaty and P. L. Patterson, Book, Comptes Rendus de la VI$^e$ Conference Internationale sur les Phenomenes D'Ionisation dans les Gaz, Ed. P. Hubert and E. Cremieu-Alcan III, pp. 289–293 (S. E. R. M. A. Publ. Co., Paris, France, 1963).

Ranking laboratories and evaluating methods of measurement in round-robin tests, T. W. Lashof, Mater. Res. Std. **4,** No. 8, 397–407 (Aug. 1964).

Projection X-ray microscopy of roofing materials, S. B. Newman and S. H. Greenfeld, Mater. Res. Std. **3,** No. 11, 893–898 (Nov. 1963).

Polymorphism of $ABO_3$-type rare earth borate solid solutions, R. S. Roth, J. L. Waring, and E. M. Levin, (Proc. Third Rare Earth Research Conf., Apr. 2–24, 1963), Book, Rare Earth Research, ch. on Structure of Rare Earth Compounds, pp. 1030–1055 (Gordon & Breach Science Publishers, Inc., New York, N.Y. 1964).

Polarography as an analytical tool, J. K. Taylor, J. Assoc. Offic. Agri. Chem. **47,** No. 1, 21–27 (Jan. 1964).

Polar-cap absorption, D. K. Bailey, Planetary Space Sci. **12,** 495–541 (May 1964).

Polymorphism in isopropyl alcohol, G. S. Ross, H. L. Dixon, L. Frolen, and D. J. Termini, Science **141,** No. 3585, 1043–1044 (Sept. 1963).

Changes in relation between refractive index and Young's modulus as the result of successive heat treatment, E. H. Hamilton, J. Am. Ceram. Soc. **47,** No. 4, 167–170 (Apr. 1964).

Characteristics of programs for KWIC and other computer-produced indexes, W. W. Youden, (Proc. American Documentation Institute Annual Meeting, Oct. 1963) Book, Automation and Scientific Communication, pp. 331–332 (H. T. Luhn, Am. Documentation Inst., 1963).

Plastics, G. M. Klein, The 1964 American Annual, p. 532 (Americana Corp. Chicago, Ill., 1964).

Oxygen atom reactions with condensed olefins, M. D. Scheer and R. Klein, Science **144,** No. 3623, 1214 (Apr. 1964).

Optically pumped magnetometers and related experiments in high magnetic fields, P. L. Bender, Book, Quantum Electronics, Ed. P. Grivet and N. Bloembergen, Ill. **1,** 263–273 (Columbia University Press, New York, N.Y., 1964).

Bound states in a Debye-Hueckel potential, C. R. Smith, Phys. Rev. **134,** A1235–A1237 (June 1964).

The units and standards of electrical measure, F. K. Harris, Electron World **72,** No. 2, 29–32 (Aug. 1964).

Tetragonal phases of the general type $10M_2O_5 \cdot 90M'_2O_5$ apparently isostructural with $Ta_2O_5 \cdot 2N_bO_5$, J. L. Waring and R. S. Roth, Acta Cryst. **17,** 455–456 (1964).

Ionizing radiation and constitution of the atmosphere, T. E. VanZandt, Proc. URSI XIVth General Assembly, Sept. 1963 Tokyo, Japan **XIII,** 158–164 (1964).

Interpretation of frequency shifts due to electron exchange collisions, P. L. Bender, Phys. Rev. **134,** No. 5A, A1174–A1180 (June 1, 1964).

International Geophysical Calendar for the years of the Quiet Sun 1964–1965, A. H. Shapley and J. V. Lincoln, J. Radio Res. Lab. Japan **10,** 185–192 (Mar. 1963); IUGG Chronicle **50,** 171–174 (July 1963); URSI Inform. Bull. **139,** 64–71 (July-Aug. 1963); IGY Bull. **74,** 15–18 (Aug. 1963); Weatherwise **16,** 160–161 (Aug. 1963); J. Geophys. Res. **68,** 5318–5323 (Sept. 1963); Trans. Am. Geophys. Union **44,** 821–824 (Sept. 1963); J. Atmospheric Terrest. Phys. **26,** No. 1, 151–157 (Jan. 1964); Nature **201,** 11–13 (Jan. 4, 1964).

International Geophysical calendar 1963, A. H. Shapley and J. V. Lincoln, IQSY Notes **4,** 34 (Sept. 1963).

Internal friction in rutile containing point defects, J. B. Wachtman, Jr., and L. R. Doyle, Phys. Rev. **135A,** 276, (1964).

Instant color correction, C. S. McCamy, Modern Photo. **28,** No. 1, 66–67 (Jan. 1964).

Review of static seals for cryogenic systems, R. F. Robbins and P. R. Ludtke, J. Spacecraft Rockets **1,** No. 2, 91 (May–June 1964).

Resurgence of solar and geomagnetic activity in September and October 1963, A. H. Shapely, IQSY Notes **5,** 10, (Dec. 1963).

Dose fields from plane sources using point source data, J. H. Hubbell, Nucleonics **21,** No. 8, 144–148 (Aug. 1963).

Diurnal and seasonal variations of the atmosphere near the 100-kilometer level, J. W. Wright, J. Geophys. Res. **69,** 2851–2853 (June 1964).

*D*-region electron density profiles during auroras, G. M. Lerfald, C. G. Little, and R. Parthasarathy, J. Geophys. Res. **69,** 2857–2860 (June 1964).

Binary error probability due to an adaptable fading model, M. Nesenbergs, IEEE Trans. Comm. Systems **CS-12,** No. 1, 64–73 (Mar. 1964).

An X-ray determination of the Debye temperature of silver iodide, G. Burley, J. Phys. Chem. Solids **25,** 629 (1964).

Hydrogen atom addition to solid four-carbon olefins, R. Klein and M. D. Scheer, J. Phys. Chem. **67,** 1874–1877 (Sept. 1963).

HF propagation characteristics in equatorial latitudes, A. Barghausen, G. Jacobs, and D. Lillie, Program & Digest of IEEE Trans. Ant. Prop. Intern. Symp. July 9–11, 1963, Boulder, Colo., pp. 37–42 (IEEE, New York, N.Y., 1963).

Hall generators: what, why and how much?, S. Rubin, 1964 IEEE Intern. Convention Record, Part 8, p. 74 (Mar. 1964).

The half-life of carbon-14: Comments on the mass-spectrometric method, E. E. Hughes and W. B. Mann, Intern. J. Appl. Rad. Isotopes **15,** No. 1, 97–100 (Jan. 1964).

A unitary symmetry selection rule and its application to new resonances, C. A. Levinson, H. J. Lipkin, and S. Meshkov, Il Nuovo Cimento **32,** Ser. X, 1376–1379 (1964).

A recalibration of the National Bureau of Standards tritiated water standards by gas counting, W. B. Mann, R. W. Medlock, and O. Yura, Intern. J. Appl. Rad. Isotopes **15,** No. 6, 351–361 (1964).

Anomalous transmission of rare gases for electrons of sub-excitation energies, J. A. Simpson (Proc. Third Intern. Conf. Physics of Electronic and Atomic Collisions, London, England, 1963), pp. 128–134 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

The guided propagation of ELF and VLF radio waves between the earth and the ionosphere, D. D. Crombie and A. G. Jean, Proc. URSI XIVth General Assembly, Sept., 1963, Tokyo, Japan, **XIII,** 195–206 (1964).

Electronic scanning for large radiotelescopes, T. A. Clark, Digest of 1963 IEEE Professional Group on Antennas and Propagation, pp. 242–246 (July 1963).

Electron microscopy and diffraction of aluminum oxide whiskers, D. J. Barber, Phil. Mag. **10,** No. 103, 75–94 (July 1964).

The excitation of the 2P state of hydrogen by electrons of near-threshold energy, G. E. Chamberlain, S. J. Smith, and D. W. O. Heddle, Phys. Rev. Letters **12,** No. 23, 647–649 (June 8, 1964).

Theory of operation of the rotary positive displacement meter, C. T. Collett (Proc. 23d Annual Gas Measurement Short Course, Aug. 20, 1963), West Virginia Univ. Tech. Bull. **70,** 48–53 (1963).

The establishment and maintenance of the unit of voltage at the NBS Boulder Laboratories, B. A. Wickoff, ISA 1963 Proc. Paper No. 28.2.63 (1963).

The crystal structure of 1-ethyldecaborane, A. Perloff, Acta Cryst. **17,** 332 (1964).

Optically observed inner shell electron excitation in neutral Kr and Xe, K. Codling and R. P. Madden, Phys. Rev. Letters **12,** No. 4, 106–107 (Jan. 1964).

On the production of the $Z^-(\Omega^-)$ Baryon, H. J. Lipkin, C. A. Levinson, and S. Meshkov, Phys. Letters **7,** No. 2, 159–161 (Nov. 1, 1963).

On the implication of diurnal, seasonal and geographical variations in composition of the high atmosphere from *F*-region measurements, J. W. Wright, J. Geophys. Res. **68,** No. 19, 5413–1520 (Oct. 1, 1963); Book, Electron Density Distribution in Ionosphere and Exosphere (Proc. of NATO Advances Study Institute, April 17–26, Sheikampen, Norway), pp. 105–115 (North Holland Publ. Co., Amsterdam, The Netherlands, 1963).

On the effect of heterogeneity in molecular weight on the sedimentation equilibrium second virial coefficient of polymers in good solvents, I. H. Billick, J. Phys. Chem. **68,** No. 7, 1798–1800 (July 1964).

Four-point probe measurement of non-uniformities in semiconductor sheet resistivity, L. H. Swartzendruber, Solid State Electronics **7,** 413–422 (1964).

The use of dual-polarized broad beam antennas to determine the extraterrestrial intensity of the cosmic radio noise at high frequencies, C. G. Little, G. M. Lerfald, and R. Parthasarathy, Digest of the 1963 IRE Professional Group on Antennas and Propagation, pp. 9–15 (July 1963).

The solid state of rare gases, G. L. Pollack, Rev. Mod. Phys. **36,** No. 3, 748–792 (July 1964).

The present and future of astrophysics and its effects on industry in Colorado, L. M. Branscomb, Book, Colorado and the New Technological Revolution: Proceedings of the University-Industry Liaison Conference 1963, Ed. W. H. Miernyk, pp. 55–64 (Bureau of Economic Research, Institute of Behavioral Science, Univ. of Colorado, Boulder, 1964).

A simplified freezing point apparatus, G. S. Ross and A. R. Glasgow, Jr., Anal. Chem. **36,** No. 3, 700 (Mar. 1964).

Autoionization in helium using a synchrotron light source, R. P. Madden and K. Codling, Proc. VIth Intern. Conf. of Ionization Phenomena in Gases, Paris, France, 1963, **1,** 139–143 (1963).

Radiowave scattering in the ionosphere, K. L. Bowles, Book, Advances in Electrons and Electron Physics **19,** 55–176 (1964).

Radio, the ionosphere, and IQSY, R. W. Knecht, Proc. URSI XIVth General Assembly, Sept. 1963, Tokyo, Japan **XIII,** 169–180 (1964).

Pseudopentagonal twins in electrodeposited copper dendrites, F. Ogburn, B. Paretzkin, and H. S. Peiser, Acta Cryst. **17,** 774–775 (1964).

Dielectric polarizability of fluid parahydrogen, J. W. Stewart, J. Chem. Phys. **40,** 3297–3306 (June 1964).

Conduction-cooled, concentric sphere glass trap, S. Ruthberg, J. Vacuum **14,** No. 1, 11 (Jan. 1964).

Conductance of potassium chloride in highly purified *n*-methylpropionamide from 20 to 40°, T. B. Hoover, J. Phys. Chem. **68,** No. 14, 876–879 (Apr. 1964).

Field measurements of air infiltration in ten electrically heated houses, C. W. Coblentz, and P. R. Achenbach, ASHRAE J. **5,** No. 7, 69 (July 1963).

Discrepancy in conversion of early melting-point temperatures, S. J. Schneider, J. Am. Ceram. Soc. **46,** No. 7, 354–355 (1963).

Relation of solar active regions at central meridian passage to ionospheric disturbance, C. S. Warwick and J. V. Lincoln, AGARDograph **59,** The Effect of Disturbances of Solar Origin on Communications, Ed. G. J. Gassmann, pp. 229–233 (Pergamon Press Inc., London, England, 1963).

Report of committee E-2 on emission spectroscopy, R. E. Michaelis, Am. Soc. Testing Mater. Proc. **63,** 456 (1963).

The relationship between intensity of carbon arc and asphalt oxidation, J. R. Wright, P. G. Campbell, and T. L. Fridinger, ASTM Spec. Tech. Publ. No. 347, pp. 3–20 (May 1964).

Evidence for field-aligned ionization irregularities between 400 and 1000 km above the earth surface, T. E. VanZandt, W. Calvert, R. W. Knecht, and G. B. Goe, Book, Space Research, Ed. W. Priester, III, 271–273 (North Holland Publ. Co., Amsterdam, The Netherlands, 1963).

Experimental predictions from unitary symmetry ($SU_3$), C. A. Levinson, H. V. Lipkin, and S. Meshkov (Proc. 1963 Intern. Conf. Nucleon Structure, Stanford Univ., Calif.), Book, Nucleon Structure, p. 309 (University Press, Stanford, Calif., Aug. 1963).

Standard potential of the Ag-AgCl electrode in 5% aqueous mannitol, R. Gary and R. A. Robinson, J. Chem. Eng. Data **9,** No. 3, 376–378 (July 1964).

Summary of symposium on recent research on bituminous materials, S. H. Greenfield, ASTM Spec. Tech. Publ. No. 347, pp. 108–111 (May 1964).

First results in NASA topside sounder satellite, R. W. Knecht, S. Russell, and J. E. Jackson, Book, Advances in Astronautical Sciences **11,** 61–85 (Western Periodicals Co., Calif., 1963).

Reaction rates for the formation of $He_2^+$ and $N_2^+$, E. C. Beaty and P. L. Patterson, Book, Comptes Rendus de la VI$^e$ Conf. Intern. sur les Phenomenes d'Ionisation dans les Gaz, Ed. P. Hubert and E. Cremieu-Alcan III, pp. 289–293 (S.E.R.M.A. Publ. Co., Paris, France, 1963).

Comparisons of experimental reaction cross sections with various relations obtained from $SU_3$, S. Meshkov, Proc. Conf. Symmetry Principles at High Energy, Coral Gables, Fla., Jan. 1964, pp. 104–126 (1964).

Very-low-frequency radio waves and the ionosphere, D. D. Crombie, Science **142,** No. 3591, 508–510 (Oct. 25, 1963).

Standardization of dimensions of packages for self-service selling, A. T. McPherson and A. S. Best, THE PACKET, Packaging Professional **2,** No. 3, 4–5 (Packaging Inst., 1964).

Response characteristics of a $4\pi$ neutron detector, H. Gerstenberg, Bull. Am. Phys. Soc. **9,** Pt. 2, 421 (Apr. 28, 1964).

Geomagnetic and solar data, J. V. Lincoln, J. Geophys. Res. **69,** No. 13, 2841 (July 1, 1964).

Spectroscopic investigations of fluorescence and chemiluminescence in gases (collection of papers), T. Carrington, Aero-Space Research Laboratory Report No. 64-51 (Mar. 1964).

Fractionation of uranium isotopes and daughter products in weathered granite and uranium-bearing sandstone, Wind River basin region, Wyoming, J. N. Rosholt, E. L. Garner, and W. R. Shields, U.S. Geol. Surv. Prof. Paper 501-B, pp. 884–887 (1964).

Some results of a new method for obtaining ionospheric N(h) profiles with a bearing on the structure of the lower *F* region, J. W. Wright and A. K. Paul, J. Geophys. Res. **68,** No. 19, 5413–5420 (Oct. 1, 1963); Proc. NATA Advanced Study Institute, Skeikampen, Norway, Apr. 17–26, 1963), Book, Electron Density Distribution in the Ionosphere and Exosphere, pp. 105–115 (North Holland Publ. Co., Amsterdam, The Netherlands, 1963).

$Sr^{87}/Sr^{86}$ ratios in Hawaiian lavas, P. Lessing and E. J. Catanzaro, J. Geophys. Res. **69,** No. 8, 1599–1601 (Apr. 15, 1964).

Optical investigations of film formation and removal on gold anodes in acidic oxalate solutions, W. E. Reid and J. Kruger, Nature **203,** No. 4943, 402 (July 25, 1964).

Opening angles of electron-positron pairs, H. Olsen, Phys. Rev. **131,** No. 1, 406–415 (July 1, 1963).

Some impulse sparkover voltages for a 200 cm sphere gap, H. N. Cones, IEEE Component Parts Conf. Paper No. CP 63-929 (June 1963).

Determination of optical thickness gradients from a far distance, J. R. Meyer-Arendt, R. B. Herrick, and C. B. Emmanuel, J. Opt. Soc. Am. **53,** 1344 (1963).

Solar particles and interplanetary fields, C. S. Warwick, Book, Advances in the Astronautical Sciences **13,** 666–681 (Western Periodicals, Calif., 1963).

Vaporization of uranium monophosphide, K. A. Gingerick, P. K. Lee, and J. E. Efimenko, Nature **200,** No. 4908, 774–776 (1963).

A radio-meteorological study, B. R. Bean, E. J. Dutton, V. R. Frank, J. A. Lane, and W. B. Sweezy, Program and Digest 1963 IEE Prof. Group Antennas and Propagation Symp. Space Telecommunications, July 9–11, 1963, Boulder, Colo., pp. 56–60 (1963).

The interaction of 350 keV polarized neutrons with oriented $Ho^{165}$ nuclei, H. Marshak, R. Wagner, P. P. Miller, and T. Tamura, Physics Letters **10,** No. 3, 316 (June 15, 1964).

Systematic atmospheric refraction errors of baseline type radio tracking systems and methods of their correction, G. D. Thayer and B. R. Bean, Proc. First Space Congress, Cocoa Beach, Florida (Apr. 1964).

Summary of session I and II of the symposium on X-ray and electron probe analysis, R. E. Michaelis, ASTM Spec. Tech. Publ. No. 349 (Dec. 1963).

Prebunching in electron linacs, J. E. Leiss, Proc. Conf. Proton Linear Accelerators, Yale Univ., New Haven, Conn., pp. 267–273 (Oct. 1963).

Report on standard samples, reference samples, and high-purity material for spectrochemical analysis, R. E. Michaelis, ASTM Spec. Tech. Publ. No. 58-E (1963).

Phase diagrams for ceramists, Book, E. M. Levin, C. R. Robbins, and H. F. McMurdie, American Ceramic Society, Inc., Columbus, Ohio (1964).

Transverse propagation of electromagnetic waves in a cylindrically stratified plasma, J. R. Wait, Proc. Radio Physics Symp., University of California, Los Angeles, Aug. 7–9, 1963, Ed. J. Brandstatter (Holden-Day, Inc., San Francisco, Calif., 1964).

Variations in phase path of man-made one-hop whistler mode signals at 18.6 kilocycles per second, D. D. Crombie, F. A. McNeill, and G. M. Allcock, J. Geophys. Res. **68,** No. 23, 6229–6235 (Dec. 1, 1963).

Recently discovered autoionizing states of krypton and xenon in the λ 380–600A, R. P. Madden and K. Codling, J. Opt. Soc. Am. **54,** No. 2, 268–269 (Feb. 1964).

Fluorometric demonstration of tryptophan in dentin and bone protein, K. C. Hoerman and S. A. Mancewicz, J. Dental Res. **43,** No. 2, 276–280 (Mar.–Apr. 1964).

The calibration of inductive voltage dividers and analysis of their operational characteristics, T. L. Zapf, ISA Trans. **2,** No. 3, 195–200 (July 1963).

Nighttime variations of *F*-region electron density profiles at Puerto Rico, T. Shimazaki, J. Geophys. Res. **69,** No. 13, 2781–2797 (July 1, 1964).

Obscurities of oscillator noise, L. Fey, W. R. Atkinson, and J. Newman, Proc. IEEE Letter **52,** 104–105 (Jan. 1964).

Standard electromotive force of the hydrogen-silver chloride cell and the thermodynamics of solutions of hydrochloric acid in 50 wt. % methanol from 10 to 40 °C, M. Paabo, R. A. Robinson, and R. G. Bates, J. Chem. Eng. Data **9,** No. 3, 374–376 (July 1964).

The *F*-region seasonal anomaly, J. W. Wright, J. Geophys. Res. **68,** No. 14, 4379–4381 (July 15, 1963).

Coherent Raman effect in the off-axis Raman resonator, H. Takuma and D. A. Jennings, Appl. Phys. Letters **4,** No. 11, 185–186 (June 1964).

Effects of composition and irradiation on the glass transition temperature of methyl methacrylate-styrene copolymers, M. S. Parker, V. J. Krasnansky, and B. G. Acchammer, J. Appl. Polymer Sci. **8,** 1825–1838 (1964).

"U-spin equalities" and octet symmetry breaking, S. Meshkov, G. A. Snow, and G. B. Yodh, Phys. Rev. Letters **13,** No. 6, 212–217 (Aug. 10, 1964).

The scope of electrochemistry, W. J. Hamer, J. Electrochem. Soc. **110,** No. 6, 115C (June 1963).

Report on sugars and sugar products, E. J. McDonalds, J. Assoc. Official Agri. Chem. **46,** 35 (1963).

Thermal conductivity and viscosity of simple fluids, J. V. Sengers, Am. Chem. Soc. **9,** No. 4, C-57–C-70 (Sept. 1964).

Some recent developments in noble metal thermocouples, L. O. Olsen, Am. Soc. Auto. Engrs. Paper No. 750A, p. 10 (Sept. 1963).

Gallium-palladium alloys as dental filling material, R. M. Waterstrat and R. W. Longton, Public Health Rept. **79,** No. 7, pp. 638–642 (July 1964).

Minutes of the meeting at the Ohio State University, Columbus, Ohio, of the Triple Commission for Spectroscopy, June 13, 1963, C. E. Moore, J. Opt. Soc. Am. **54,** No. 3, 425–432 (Mar. 1964).

High frequency microvolt measurements, F. X. Ries, ISA Preprint, pp. 37.2.63-1 – 37.2.63-5 (Sept. 1963).

Comparison of a new $SU_3$ prediction with experiment, S. Meshkov, G. A. Snow, and G. B. Yodh, Phys. Rev. Letters **12,** No. 3, 87–92 (Jan. 20, 1964).

Measurement of photochemical degradation in certain plastics by color reactions with phenylenediamines, V. E. Gray and J. R. Wright, J. Appl. Polymer Sci. **8,** No. 4, 1505–1519 (1964).

On Rayleigh's non-linear vibration equation, A. Ghaffari, Proc. Intern. Symp. Non-Linear Vibration II, 130–133 (Sept. 1961).

Electrode potentials in fused systems. VII. Effect of ion size on membrane potentials, K. H. Stern and J. A. Stiff, J. Electrochem. Soc. **III,** No. 8, 893–897 (Aug. 1964).

An unmodulated twin channel microwave attenuation measurement system, D. H. Russell, ISA Preprint, pp. 37.3.63-1–37.3.63-17 (Sept. 1963).

Beam loading and beam blowup in electron linacs, J. E. Leiss, Proc. Conf. Proton Linear Accelerators, Yale Univ., New Haven, Conn., pp. 74–95 (Oct. 1963).

Residual range and range rate errors due to the troposphere, E. C. Barrows, Proc. Conf. Tropospheric Refraction (Dec. 1963).

Tropospheric refraction and attenuation of radio waves, B. R. Bean, Book, Advances of Radio Research, pp. 53–155 (Academic Press, Inc., New York, N.Y., June 1964).

Kikuchi pattern from a silicon wedge, H. A. Fowler, Acta Cryst. **17,** No. 7, 869–871 (July 1964).

On line data handling system for the National Bureau of Standards LINAC, J. Leiss, J. Broberg, R. Schrack, and J. M. Wyckoff, IEEE Trans. Nuclear Sci. **NS-11,** No. 3, 331–337 (June 1964).

Long-term characteristic for air-ground propagation in band nine, R. S. Kirby, Program and Digest 1963 IEEE Prof. Group Antennas and Propagation Symp. Space Telecommunications, July 9–11, 1963, Boulder, Colo., p. 43–47 (1963).

Measurement of effective temperatures of microwave noise sources, J. S. Wells, W. C. Daywitt, and C. K. S. Miller, IEEE Trans. Instr. Meas. **IM-13,** No. 1, 17–28 (Mar. 1964).

The turbulent characteristics of the radio refractive index near the ground, B. R. Bean and R. E. McGavin, Program and Digest 1963 IEE Prof. Group Antennas and Propagation Symp. Telecommunications, July 9–11, 1963, Boulder, Colo., pp. 60–64 (1963).

Use of high-strength ceramics in vibration transducers, E. Jones, J. Acoust. Soc. Am. **36,** No. 6, 1215–1216 (June 1964).

Thermodynamics of aqueous solutions of hydriodic acid from electromotive force measurements of hydrogen-silver iodide cells, H. B. Hetzer, R. A. Robinson, and R. G. Bates, J. Phys. Chem. **68,** No. 7, 1929–1933 (1964).

Phase instability in a tropospheric relay link, E. C. Barrows, Program and Digest 1963 IEEE Prof. Group. Antennas and Propagation Symp. Space Telecommunications, July 9–11, 1963, Boulder, Colo., pp. 52–55 (1963).

Analysis of rotation errors of a waveguide rotary-vane attenuator, W. Larson, IEEE Trans. Instr. Meas. **IM-12,** No. 2, 50–55 (Sept. 1963).

The radio thermal noise properties of the lower atmosphere, B. R. Bean, Program and Digest 1963 IEEE Prof. Group. Antennas and Propagation Symp. Space Telecommunications, July 9–11, 1963, Boulder, Colo., pp. 30–31 (1963).

The oxide films formed on copper single cyrstal surfaces in water. III. Effect of light, J. Kruger and J. P. Calvert, J. Electrochem. Soc. **III,** No. 9, 1038–1041 (Sept. 1964).

Weights and measures – a challenge to industry, M. W. Jensen, Proc. 50th Mid-Year Meeting Chemical Specialties Manufacturers Assoc., Chicago, Ill., May 18–20, 1964, pp. 24–26 (Chem. Specialties Mfg. Assoc. Inc., New York, N.Y., 1964).

Interpretation of potentiometric titrations of weak acids in methanol-water solvents, K. C. Ong, R. A. Robinson, and R. G. Bates, Anal. Chem. **36,** 1971–1972 (Sept. 1964).

Fluctuations in ionospheric absorption events at conjugate stations, J. K. Hargreaves and H. J. A. Chivers, Nature **203,** No. 4948, 963–964 (Aug. 29, 1964).

RF and microwave power measurements, G. F. Engen, Proc. URSI XIVth General Assembly, Tokyo, Japan, Sept. 1963, **XIII,** 77–80 (1964).

LF–VLF frequency and time services of the National Bureau of Standards, D. H. Andrews (Proc. 19th Annual ISA Conf. and Exhibit, New York, N.Y., Oct. 12–15, 1964), ISA Preprint 21.3–2–64 (1964).

Fresnel zone diffraction effects at 50 Gc/sec. determined from measured aperture field data, R. C. Baird, Program and Digest 1963 IEEE Prof. Group Ant. and Prop. Symp. Space Telecommunications, July 9–11, 1963, Boulder, Colo., pp. 171–173 (1963).

Computer interpretation of English text and picture patterns, R. A. Kirsch, IEEE Trans. Electron. Computers **EC-13,** No. 4, 363–376 (Aug. 1964).

Electronic energy bands in strontium titanate, A. H. Kahn and A. J. Leyendecker, Phys. Rev. **135,** No. 5A, 1321–1325 (Aug. 31, 1964).

On the implication of diurnal, seasonal and geographical variations in composition of high atmosphere from *F*-region measurements, J. W. Wright, (Proc. NATO Advanced Study Institute, Sheikampen, Norway, Apr. 17–26, 1963), Book, Electron Density Distribution in the Ionosphere and Exosphere, pp. 186–198 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

Ionospheric winds: motions into night and sporadic *E* correlations, N. W. Rosenberg, H. D. Edwards, and J. W. Wright, (Proc. 4th Intern. Space Science Symp., Warsaw, June 4–10, 1963), Book, Space Research, Ed. P. Muller, IV, 171–181 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

Oblique incidence pulse measurement at 100 kc, R. H. Dougherty, AGARDograph **74,** Propagation of Radio Waves at Frequencies below 300 KC, pp. 133–148 (Pergamon Press Inc., New York, N.Y., 1963).

Nonresonant absorption and collision diameters in the foreign-gas broadening of symmetric top molecules, G. Birnbaum and A. A. Maryott, J. Chem. Phys. **41,** No. 1, 154–157 (July 1, 1964).

Methods of calibration of airglow photometers at Fritz Peak Observatory, L. L. Smith and R. B. Alexander, IQSY Instruction Manual No. 5: Airglow, Pt. II, p. 25–36 (IQSY Secretariat, London, England, 1963).

Electromagnetic cross sections for electron and nuclear research, H. W. Koch, Nucl. Instr. Methods **28,** 199–204 (1964).

Exclusion of parity unfavored transitions in forward scattering collisions, U. Fano. Phys. Rev. **135,** No. 3B, B863–B864 (Aug. 10, 1964).

Microscopic-astrophysics and the development of astrophysical-laboratories, R. N. Thomas, Bull. Centre Intern. d'Astrophysique de l'Observatoire de Nice, No. 1, pp. 7–16 (Nice, France, 1964).

Alouette ionic, magnetic-field and plasma studies, W. Calvert, H. Rishbeth, and T. E. VanZandt, IG Bull. No. 83, 16–19 (May 1964); Trans. AGU **45,** No. 2, 398–401 (June 1964).

Ozonation of asphalt flux, P. G. Campbell and J. R. Wright, I. & E. C. Prod. Res. and Devel. **3,** 186–194 (Sept. 1964).

Propagation of radio waves past a coast line with a gradual change of surface impedance, J. R. Wait and K. P. Spies, IEEE Trans. Ant. Prop. **AP-12,** No. 5, 570–575 (Sept. 1964).

On the propagation of long wave-length terrestrial radio waves – two theoretical techniques, J. R. Johler and L. A. Berry, AGARDograph **74,** Propagation of Radio Waves below 300 KC, pp. 101–129 (Pergamon Press Inc., New York, N.Y., 1963).

On the validity of the Einstein-Smoluchowski theory light scattering, R. Zwanzig, J. Am. Chem. Soc. **86,** 3489–3493 (Mar. 30, 1964).

Experiments on self-ionizing shock waves in a magnetic field, Y. Nakagawa and K. B. Earnshaw, Book, Comptes Rendus de la VI$^e$ Conference Internationale sur les Phenomenes D'Ionisation dans les Gaz, ed. P. Hubert and E. Cremieu-Alcan, IV, 303–309 (S.E.R.M.A., Paris, France, 1963).

Microwave propagation in an overdense bounded magnetoplasma, B. Wieder, Phys. Fluids **7,** 964–972 (July 1964).

Pulsed and CW sinusoidal voltage and current measurements, M. C. Selby, Proc. URSI XIVth General Assembly, Tokyo, Japan, Sept. 1963, **XIII,** 92–100 (1964).

Thermodynamic data compilation and review at the National Bureau of Standards, G. T. Armstrong, Svensk Kem. Tidskr. **76,** No. 8, 419–427 (1964).

Ionosonde studies of some chemical releases in the ionosphere, J. J. W. Wright, (Proc. NATO Advanced Study Institute, Sheikampen, Norway, Apr. 17–26, 1963), Book, Electron Density Distribution in the Ionosphere and Exosphere, pp. 116–137 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

Standard nuclear instrument modules, L. Costrell, U.S. Atomic Energy Commission TID Report 20893, 12 pages (July 1964).

Charge-transfer absorption spectra of NO in Kr and $CH_3OH$ solutions, E. E. Ferguson and H. P. Broida, J. Chem. Phys. **40,** No. 12, 3715–3716 (June 15, 1964).

Ionospheric storms, S. Matsushita, Proc. URSI XIVth General Assembly, Tokyo, Japan, Sept. 1963, **XIII,** 169–180 (1964).

Symplectic modulary groups, M. Newman and J. R. Smart, Acta Arithmetica **IX,** 83–89 (1964).

Propagation of a reflected shock wave in a transverse magnetic field, Y. Boublil and K. B. Earnshaw, Phys. Fluids **7,** No. 6, 876–881 (June 1964).

Refractivities of the rare earth oxides, E. M. Levin, (Proc. Third Rare Earth Research Conf., Apr. 2–24, 1963), Book, Rare Earth Research, Chapter on Magnetic and Optical Studies, pp. 339–350 (Gordon and Breach, Science Publ., Inc., New York, N.Y., May 1964).

Standards and measurement of attenuation, impedance and phase shift, R. W. Beatty, Proc. URSI XIVth General Assembly, Tokyo, Japan, Sept. 1963, **XIII,** 80–92 (1964).

Vapour pressure, G. T. Armstrong, Encyclopaedic Dictionary of Physics **7,** 586–587 (Pergamon Press Inc., New York, N.Y., 1962).

Analysis of a differential phase shifter, D. A. Ellerbruch, IEEE Trans. Microwave Theory Tech. **MTT–12,** No. 4, 453–459 (July 1964).

Exterior corrosion of cast-iron pipe, M. Romanoff, J. Am. Water Works Assoc. **56,** No. 9, 1129–1143 (Sept. 1964).

The energy loss of electrons in solids, A. M. Rauth and J. A. Simpson, Radiation Res. **22,** No. 4, 643–661 (Aug. 1964).

Coherence theories of tropospheric radio propagation, J. R. Wait, IEEE Trans. Ant. Prop. **AP–12,** No. 5, 649–651 (Sept. 1964).

Solar activity, J. V. Lincoln, IQSY Instruction Manual. No. 2, Report of sudden ionospheric disturbances, Suppl. pp. 1–11 (IQSY Secretariat, London, England, 1963).

Comparison calibration of inductive voltage dividers, R. V. Lisle and T. L. Zapf, ISA Trans. **3,** No. 3, 238–242 (July 1964).

Vibration-rotation bands of deuterium cyanide and hydrogen cyanide, A. G. Maki, E. K. Plyler, and R. Thibault, J. Opt. Soc. Am. **54,** No. 7, 869–876 (July 1964).

IQSY Instruction Manual. No. 1. World days, A. H. Shapley, 32 pages, IQSY Secretariat, London, England (July 1963).

Heater lead problem in calorimetry, D. C. Ginnings and E. D. West, Rev. Sci. Instr. **35,** No. 8, 965–967 (Aug. 1964).

Aspects of the National Bureau of Standards instrumentation literature reference file, J. F. Mayo-Wells, (ISA 19th Annual Conf. and Exhibit, New York, N.Y., Oct. 12–15, 1964), ISA Preprint 20.6–5–64 (1964).

Influence of the refractive index profile in VHF reflection from a tropospheric layer, J. R. Wait and C. M. Jackson, IEEE Trans. Ant. Prop. **AP–12,** No. 4, 512–513 (July 1964).

Photometric observations of the airglow during the IQSY, F. E. Roach, IQSY Instruction Manual. No. 5: Airglow, Pt. I, 1–24 (IQSY Secretariat, London, England, 1963).

Correlation matrix for radio waves in the troposphere, E. C. Barrows, (Proc. Symp. Electromagnetic Theory and Antennas, Copenhagen, 1962), Monograph, Electromagnetic Theory and Antennas 2, Sec. C, 743–746 (Macmillan and Co., Ltd., London, England, 1963).

Frequency and time broadcast services of the National Bureau of Standards, A. H. Morgan, Proc. ISA 17th Annual Automation Conf. and Exhibit, New York, N.Y., Oct. 15–18, 1962, Pt. I, Paper 1.1.62, 9 pages (1962).

Frostbite: signs, prevention, and treatment, J. R. Meyer-Arendt, Trail and Timberline, p. 3 (Jan. 1964).

Formation of an electron depleted region in the ionosphere by chemical releases, D. Golomb, N. W. Rosenberg, J. W. Wright, and R. A. Barnes, (Proc. 4th Intern. Space Science Symp., Warsaw, June 4–10, 1963), Book, Space Research Ed. P. Muller, IV, 389–398 (North Holland Publ. Co., Amsterdam, The Netherlands, 1964).

The ionosphere, A. H. Shapley, U.S. Information Agency Voice of America Forum Series on Space Science, Lecture No. 7, 6 pages (1963).

Infrared spectroscopy of weak charge-transfer complexes, E. E. Ferguson, J. Chim. Phys. **61,** 257–262 (1964).

**Publications for which a price is indicated are available by purchase from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402 (foreign postage, one-fourth additional). Reprints from outside journals and the NBS Journal of Research may often be obtained directly from the authors.*

## Horace M. Trent

Dr. Horace M. Trent, Head of the Applied Mathematics Staff, Naval Research Laboratory, died December 16, 1964, following an illness complicated by pneumonia. He was born in Chattanooga, Tenn., in 1907, and educated at Berea College, where he received his B.A. in 1928. In 1929 he received his M.A. and in 1934 was awarded his doctoral degree trom Indiana University. He then served for a number of years as a professor of physics, magnetism, acoustics, and electronics at Mississippi A&M College and Mississippi State University. In 1940 he came to the U.S. Naval Research Laboratory and at the time of his death was Head of the Applied Mathematics Staff there.

Dr. Trent did research in acoustics, mechanics, thermodynamics, and electronics. This work resulted in many papers and patents. His recent interests were in linear graph theory and its applications to electromechanical analogies.

Dr. Trent was a member of and held responsible positions in numerous professional and social organizations. He was a Fellow of the Acoustical Society of America, American Association for the Advancement of Science, and the Washington Academy of Sciences; he was a member of the American Mathematical Society, Society for Experimental Stress Analysis, Philosophical Society of Washington, Society of Engineering Science, Society for Industrial and Applied Mathematics, and the American Standards Association. He was also a member of the Cosmos Club and the Palaver Club.

U.S. GOVERNMENT PRINTING OFFICE : 1965 OL—757-615

SJR-8 AUG
UNIVERSITY MICROFILMS, INC.
ANN ARBOR, MICHIGAN

UNITED STATES
GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
WASHINGTON, D.C. 20402

OFFICIAL BUSINESS

PENALTY FOR PRIVATE USE TO AVOID
PAYMENT OF POSTAGE, $300
(GPO)

JOURNAL OF RESEARCH of the National Bureau of Standards — Vol. 69B, No. 1 — January–March 1965
B. Mathematics and Mathematical Physics — No. 2 — April–June 1965

## Contents

| | Page |
|---|---|
| Lectures on matroids. W. T. Tutte. | 1 |
| Menger's theorem for matroids. W. T. Tutte. | 49 |
| Single-element extensions of matroids. Henry H. Crapo. | 55 |
| Minimum partition of a matroid into independent subsets. Jack Edmonds. | 67 |
| Lehman's switching game and a theorem of Tutte and Nash-Williams. Jack Edmonds. | 73 |
| On the connection between the properties of oriented linear graphs and analyses of lumped physical systems. Horace M. Trent. | 79 |
| Character subgroups of F-groups. Marvin I. Knopp and Morris Newman. | 85 |
| A note on multipliers of difference sets. R. A. Brualdi. | 87 |
| Modification of Edmonds' maximum matching algorithm. C. Witzgall and C. T. Zahn, Jr. | 91 |
| On measurable sets and functions. A. J. Goldman. | 99 |
| Characteristic relations for nonperiodic solutions of Mathieu's equation. T. Tamir and H. C. Wang. | 101 |
| On the surface duality of linear graphs. Jack Edmonds. | 121 |
| Maximum matching and a polyhedron with 0,1-vertices. Jack Edmonds. | 125 |
| Publications of the National Bureau of Standards. | 131 |

Library of Congress Catalog Card Number: 62–4414

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402. Price 75 cents (single copy). Subscription price: $2.25 a year, $0.50 additional for foreign mailing.

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1965